

THESE

présentée à

L'UNIVERSITE PIERRE ET MARIE CURIE (PARIS VI)

pour l'obtention du

DOCTORAT TROISIEME CYCLE

Spécialité: MATHEMATIQUES

Mention: INFORMATIQUE

par Madame Brigitte GRAU

Sujet de la thèse: ANALYSE ET REPRESENTATION D'UN TEXTE
D'APRES LE THEME DU DISCOURS

Soutenue le 15 décembre 1983 devant la Commission

composée de: Président M. J.C. SIMON

Examineurs: MM. H. GALLAIRE
J.M. PIERREL
G. SABAH
J. VIRBEL

Invites:

THESE

présentée à

L' UNIVERSITE PIERRE ET MARIE CURIE (PARIS VI)

pour l'obtention du

DOCTORAT TROISIEME CYCLE

Spécialité : MATHEMATIQUES

Mention : INFORMATIQUE

par Madame Brigitte GRAU

Sujet de la thèse : ANALYSE ET REPRESENTATION D'UN TEXTE
D'APRES LE THEME DU DISCOURS

Soutenue le 15 décembre 1983 devant la Commission

composée de: Président M. J.C. SIMON

Examineurs: MM. H. GALLAIRE
J.M. PIERREL
G. SABAH
J. VIRBEL

Invites :

Je remercie Monsieur le Professeur Jean-Claude SIMON d'avoir accepté que je prépare cette thèse dans son équipe. Je le remercie également de l'intérêt qu'il a porté à mes travaux.

Je tiens à remercier Monsieur Gérard SABAH de m'avoir orientée vers ce domaine de recherche. Je le remercie également pour ses conseils et ses remarques qui m'ont toujours été très utiles.

J'exprime ma gratitude à Messieurs Hervé GALLAIRE, Jean-Marie PIERREL et Jacques VIRBEL d'avoir accepté d'être examinateurs dans ce jury. Je leur suis reconnaissante de l'intérêt qu'ils ont ainsi montré pour mes recherches.

Tous les membres de l'équipe de Gérard SABAH m'ont toujours aidée au travers de discussions et de remarques pertinentes. Je les en remercie vivement.

Enfin, je tiens à remercier tous ceux qui ont permis l'achèvement de cette thèse, spécialement l'équipe de l'imprimerie de l'Institut de Programmation.

- INTRODUCTION

- CHAPITRE I: LE THEME DU DISCOURS.

I.1- Présentation des travaux traitant du sujet.	p. 14
I.1.1- Système réalisé par B.Grosz.	p. 14
I.1.2- Système réalisé par C.Sidner.	p. 17
I.1.3- Conclusion.	p. 19
I.2- Sélection du thème du discours.	p. 21
I.3- Cohérence et références.	p. 24
I.4- Déroulement du thème.	p. 26
I.5- Le contexte.	p. 32
I.6- Analyse montante et descendante.	p. 34
I.7- Représentation du discours.	p. 37
I.8- Ambiguïtés étudiées par le biais du thème.	p. 39

- CHAPITRE II: REPRESENTATION DES CONNAISSANCES.

II.1- Représentation des concepts.	p. 42
II.2- Représentation interne des phrases.	p. 45
II.3- Les schémas.	p. 47
II.3.1- Description d'un schéma.	p. 47
1) Connaissances représentées dans un schéma.	p. 47
2) Structure d'un schéma.	p. 50
3) Opérations liées à un schéma.	p. 54
II.3.2- Organisation des schémas.	p. 59
1) La structure et sa signification.	p. 59
2) Quelques points sur l'utilisation de la base de connaissances.	p. 65
II.3.3- Comparaison avec des structures de données existantes.	p. 67

- CHAPITRE III: DESCRIPTION DES PRINCIPALES PROCEDURES.

III.1- Recherche du thème principal.	p. 79
III.2- Rapport phrase-schéma et vérification de la cohérence.	p. 88
III.3- Représentation du texte et du déroulement du thème.	p. 94
III.3.1- Représentation du texte.	p. 94
III.3.2- Représentation du suivi du thème et mise à jour du thème actif.	p. 95
III.4- Construction du contexte.	p. 98
III.5- Résolution d'un type de métaphore.	p. 100
III.6- Choix du sens d'un mot en cas d'ambiguïté.	p. 103

- CHAPITRE IV: EXPLICATION D'UN EXEMPLE. p. 109

- CONCLUSION. p. 125

- ANNEXES p. 129

- BIBLIOGRAPHIE p. 134

Introduction

Le traitement du langage naturel constitue un thème de recherche depuis quelque trente années. Mais il est certains points qui constituent toujours des sujets ouverts, comme la compréhension des textes.

Les principaux travaux traitant de ce thème, en Intelligence Artificielle, constituent deux courants.

Le premier consiste à analyser entièrement un texte, en faisant intervenir un large éventail de connaissances et de processus.

Les systèmes développés par Charniak (12)(13) et Schank (40) entrent dans cette catégorie. Charniak cherche à comprendre des histoires simples, histoires d'enfants ou de peinture, et à répondre à des questions sur ces histoires. Il part d'une représentation interne des phrases.

Schank a lui aussi des buts similaires à propos d'histoires au sujet de repas au restaurant.

Pour comprendre les phrases, les raisons de chaque acte, les motivations des personnages, ils ont montré la nécessité de posséder un grand nombre de connaissances de natures différentes et en ont proposé des modèles de représentation.

Ils ont réalisé des systèmes qui utilisent toutes ces connaissances et essayent de comprendre des textes. Mais ces systèmes tendent à escamoter certains points, propres à l'analyse de texte. C'est ainsi qu'ils n'ont pas développé des sujets tels que le traitement du temps, du thème de l'histoire, des personnages ...

De plus les domaines d'étude étant limités à cause du nombre de connaissances à fournir, les systèmes réalisés ne sont pas forcément généralisables.

Les derniers travaux situés dans cette ligne sont ceux de Lehnert (32) avec la réalisation d'un système intégré comprenant des histoires de divorce.

La seconde tendance consiste à décomposer l'analyse d'un texte en études différentes, traitant chacune d'un problème spécifique.

C'est ainsi qu'ont été développés des analyseurs de phrases, Rady (36), des systèmes traitant les anaphores à l'aide du thème du discours, Grosz (26)(27), Sidner (42)(43)(44), des modèles de représentation des connaissances sémantiques ou pragmatiques, Webber (48), le traitement des personnages, Berthelin (4) ...

Cette approche permet d'approfondir le traitement d'un sujet relatif à la compréhension d'un texte et de lui conférer un caractère plus général.

Notre travail se situe dans cette optique. Mais cette méthode pose le problème de la réunion des

différents modules. Faut-il les intégrer ou garder des modules séparés ? Dans ce dernier cas, comment se font les interactions entre les différents modules ?

Ces problèmes ne sont envisagés que dans les travaux de Rady (36) et Arens (2), ce dernier décrivant un analyseur composé de deux modules qui sont en interactions : un module axé sur le langage et l'autre sur le contexte.

Rady, dans sa thèse (p.49-53), penche pour une intégration des modules afin d'éliminer le non-déterminisme persistant dans une architecture horizontale : les modules y sont en interaction.

Nous n'envisageons pas le cas d'une architecture où les modules sont tous indépendants et s'exécutent les uns après les autres. Cela correspond à une séparation trop rigide des tâches.

Il est certain qu'une architecture horizontale est difficile à réaliser car il faut limiter les échanges entre les modules. C'est pourquoi nous préférons à cela un système intégré.

Parmi les sujets liés à la compréhension d'un texte, nous avons choisi de traiter un aspect fort peu étudié jusqu'à présent: la recherche et le suivi du thème ainsi que la vérification de la cohérence du texte.

Ces problèmes ont fait l'objet de quelques travaux en Intelligence Artificielle, notamment le suivi du

thème. Les principes de cohérence d'un discours ont surtout été étudiés de façon théorique, Hobbs (30), Reichman (38). En effet, les systèmes informatiques utilisent la cohérence des textes qu'ils analysent, mais ne travaillent pas directement à sa vérification.

Pourtant, si la cohérence est nécessaire à la compréhension d'un texte, il se peut que cette condition ne soit pas toujours respectée, du moins par rapport à la personne réceptrice; cela peut provenir de connaissances ou de raisonnements différents entre l'émetteur et le récepteur d'un discours.

Or, se poser le problème sous cette forme amène à considérer la compréhension d'un texte, donc des phrases qui le composent, sous un autre point de vue.

En effet si la première des tâches à effectuer est une analyse syntaxico-sémantique des phrases, il faut ensuite vérifier la cohérence entre la phrase qui vient d'être analysée et la partie du texte déjà traitée. C'est en cela que nous différons des autres analyses; car si tout le monde s'accorde sur le fait que le thème du discours et les principes de cohérence constituent des guides importants lors de l'analyse d'un texte, personne, à notre connaissance, n'a vérifié dans un système informatique le respect de ces principes. Ils ont toujours été admis comme tel.

Les relations de cohérence sont établies par les inférences produites entre deux phrases afin de les relier. Notre travail consiste à vérifier que ces relations existent.

Nous n'avons pas cherché à suivre une théorie de la cohérence, identifiant chaque type de lien et l'expliquant en détail. Ceci n'entre pas dans le cadre de notre recherche. Pour une analyse détaillée du texte, pour découvrir les plans des locuteurs ... Il faut mettre en place d'autres méthodes d'analyse.

La vérification de la cohérence telle que nous l'envisageons doit nous permettre de suivre le développement d'un sujet, et c'est sous ce point de vue que nous abordons l'analyse de texte.

Mettre en évidence la cohérence d'un discours consiste donc à établir le lien entre les phrases qui le composent.

Ex 1: P1- Je veux organiser une réunion avec Dave.

P2- Appelle Bruce et cherche une salle.

Bien que la seconde phrase ne mentionne pas explicitement la réunion, on trouve une relation entre P1 et P2 qui rend l'ensemble cohérent.

Par contre un discours tel que Ex 2 ne sera généralement pas compris, faute de trouver un lien entre la signification des deux énoncés.

Ex 2: P1- Je veux organiser une réunion avec Dave.

P2- J'aime les bonbons à la menthe.

Nous ne cherchons pas à trouver un contexte propre à chaque énoncé et indépendant de ce qui précède. Notre but est d'intégrer chaque phrase dans un contexte défini par les phrases déjà analysées.

En quoi consiste cette opération ?

Nous soutenons qu'il est impossible de connaître le sens d'une phrase sans la relier à un contexte. En effet lorsque l'on nous présente une phrase isolée, nous créons immédiatement un contexte pour l'interpréter.

Quelle est l'utilité du contexte ? Plusieurs thèses ont cours chez les psycholinguistes.

Selon Le Ny (33), le contexte permet de souligner certains traits attachés aux concepts de la phrase; le sens de la phrase est alors la composition de ces traits.

Selon Bransford et autres (11), le contexte permet de préciser de nombreux éléments qui ne sont pas mentionnés explicitement et qui ne font pas partie de la signification lexicale des mots utilisés. Les exemples qu'ils citent sont les suivants.

Ils proposent deux énoncés aux sujets de l'expérience:

- 1) La fille a fait des bonbons.
- 2) La fille a acheté des bonbons.

Ils constatent que 2) est plus efficacement rappelé quand ils donnent l'indice "dans un magasin" que ne

l'est 1), et inversement avec l'indice "dans une cuisine".

L'information concernant un lieu, par exemple "cuisine", vient nécessairement du contexte, puisqu'elle ne constitue certainement pas un trait de "fille", ni de "faire", ni de "bonbon".

Le problème se pose alors de savoir où représenter ces inférences.

Au niveau de la compréhension des phrases, les inférences de ce type permettent de relier les énoncés d'un texte entre eux.

Le rôle du contexte peut être résumé ainsi:

- sélectionner le sens des mots de la phrase;
- relier les phrases entre elles par la production d'inférences et ainsi vérifier leur cohérence.

La plupart des psycholinguistes reconnaissent l'utilité du contexte. Mais bien peu en donnent une définition précise ou indiquent ce qu'il représente exactement. Pour cela nous nous sommes reporté à des travaux sur le thème du discours en Intelligence Artificielle.

Les liens de cohérence sont en fait les enchaînements possibles entre les sujets du texte. Le contexte dépend donc des thèmes traités. C'est par le biais de la représentation des sujets que nous avons pu définir le

contexte.

Représenter des sujets de discussion implique la description de tout ce que l'on sait sur ces sujets. Cela constitue une masse importante de connaissances.

Afin de représenter les sujets possibles et les liens qui existent entre eux, nous avons adopté la solution des schémas ("frame" pour Charniak, Goldstein et Roberts ..., "script" pour Schank). Nous avons adapté ce type de représentation à l'utilisation que nous voulons en faire, en essayant d'éviter les inconvénients qui leur sont imputés:

- structures trop rigides;
- sélection du bon schéma.

Notre travail développe donc trois idées principales:

- recherche et suivi du thème d'un discours;
- vérification de la cohérence;
- représentation des connaissances propres à chaque sujet.

Nous ne prétendons pas pouvoir analyser tous les textes grâce à cette méthode. Nous ne traitons pas les résumés, car ceux-ci n'indiquent que les grandes idées d'un texte et n'explicitent pas les relations entre chacune d'elles. Il en est de même pour les textes dont la cohérence repose essentiellement sur la connaissance des personnages, sur leurs motivations personnelles; notre programme ne possède pas de connaissances à ce sujet.

Dans le premier chapitre, après avoir analysé les systèmes de Grosz et Sidner, nous exposons les étapes de l'analyse d'un texte:

- recherche du thème principal;
- définition des types d'enchaînements des sujets;
- analyse des phrases.

Nous y précisons également le mode de représentation du discours et du thème, et quelques cas d'ambiguïtés résolues grâce au thème.

Le deuxième chapitre est entièrement consacré à la représentation des connaissances qui nous sont nécessaires. Nous y indiquons la représentation du sens des mots par un réseau sémantique et la représentation interne des phrases analysées sous forme d'arborescence.

Nous y développons surtout la représentation des thèmes au moyen d'un graphe de schémas. Nous verrons alors les principales caractéristiques de cette représentation et son originalité par rapport aux modèles existants.

Nous décrivons ensuite en détail les principales procédures de notre système (chapitre III), aussi bien celles qui concernent la recherche et le suivi du thème que celles qui déterminent le contexte ou permettent de résoudre les cas d'ambiguïtés.

Enfin le dernier chapitre sera consacré à l'explication d'un exemple traité par notre système.

Chapitre I

Le thème du discours

I.1- Présentation de travaux traitant du thème d'un discours.

I.1.1- Système réalisé par B. Grosz.

La notion de thème de discours, en intelligence artificielle, a été introduite par B. Grosz (26)(27). Dans sa thèse, elle se sert de cette approche afin de résoudre les problèmes de références et d'ellipses dans un dialogue.

Le but de son travail est la compréhension de conversations entre un expert et un apprenti qui tente de mener à bien une tâche, par exemple monter une pompe à air comprimé.

La finalité de ce travail est de pouvoir substituer une machine à un expert pour guider un apprenti.

Pour résoudre les problèmes de références et d'ellipses, B. Grosz utilise le thème du discours. Elle se propose de suivre son déroulement.

Comme elle travaille dans un cadre précis, elle n'analyse que des dialogues dont elle connaît la

structure. Cette structure découle de la nature de la tâche à accomplir. En effet la tâche initiale est décomposée en sous-tâches.

Le dialogue analysé permettant d'aider un apprenti dans l'accomplissement de ces tâches et sous-tâches, son déroulement est parallèle à celui de la tâche et la résolution des problèmes de suivi du thème et de changements de thème en sont facilités.

B. Grosz ayant a priori connaissance du thème général ("topic") et de la structure générale du discours, le processus mis en place pour suivre le déroulement des dialogues est tout à fait descendant.

Elle travaille sur des réseaux partitionnés pour représenter le thème. A chaque phrase, elle construit un ensemble de sujets possibles. Ces thèmes proviennent des concepts apparaissant dans la phrase; ce sont les thèmes explicites ("explicit focus"). Elle possède également la notion de thèmes implicites ("implicit focus"); ce sont les concepts du réseau reliés aux thèmes explicites.

A cela, s'ajoute la connaissance du thème actif; ainsi B. Grosz peut savoir où elle en est dans un dialogue.

Elle possède aussi tout un processus pour reconnaître qu'un thème est terminé - on a fini d'en parler - et que l'on passe à autre chose; ce processus repose sur la nature des phrases émises pour clore un sujet, tel que les expressions : bon, passons à la suite ...

Tout cela est possible à cause de la nature des dialogues étudiés. Dans un texte dont on ne connaît pas le sujet, et encore moins la structure, il serait inadéquat de chercher à deviner à chaque pas, le thème qui va être développé.

Cela signifierait que l'on connaît tous les enchaînements possibles, toutes les histoires.

Nous ne pouvons donc pas en déduire une méthode appropriée au suivi d'un thème. De plus, le fait d'introduire un concept dans une phrase ne signifie pas qu'il va jouer un rôle important et constituer un sujet possible. Il peut servir uniquement à situer un événement, un personnage.

C. P1: Monsieur, j'ai un canot automobile dans le port et mon moteur est en panne.

P2: Une espèce de pêcheur-matelot-joueur-de-boules sur le quai, m'a dit de m'adresser à vous.

M. P3: C'est bien gentil de sa part, mais je ne sais pas si je peux y faire grand-chose.

Exemple I.1: Extrait de César de M. Pagnol.

Dans la phrase P1, il est fait mention du port. Or ce concept ne sera jamais repris par la suite en tant que sujet de discussion. Mais cela nous permet de situer

l'action.

De même, dans la phrase P2, un nouveau personnage est introduit; mais il ne constitue pas un des personnages de l'histoire. On n'en reparlera jamais.

Ces deux exemples servent à montrer que chercher à prévoir un sujet en relevant tous les concepts de la phrase est une démarche inappropriée, à la limite de l'incohérence.

Reprenons la phrase P2. Les thèmes possibles seraient ici le pêcheur et le quai, en plus des thèmes déjà traités. Mais on serait très surpris si le texte traitait du quai, de sa taille, etc ... De plus, on peut toujours introduire un sujet sans prévenir dans la phrase précédente.

C'est pourquoi la méthode que B. Grosz a mis au point pour suivre le thème d'un discours est difficilement généralisable à des textes dont on ne connaît ni la structure ni le sujet.

Il est certain par contre que les problèmes de suivi du thème et de recherche de référents des groupes nominaux définis sont très liés.

D'autres ont travaillé aussi sur ces problèmes, ainsi, C. Sidner sur la résolution d'anaphores.

I.1.2- Système réalisé par C. Sidner.

C. Sidner (10)(42)(43)(44), a développé un module de PAL, "the PAL personal assistant program", interpré-

tant des textes où il est question d'emplois du temps.

Elle s'appuie sur une représentation des connaissances sous forme de "frames" (que nous traduisons par schéma) du langage FRL, Goldstein (22).

C. Sidner, même si elle ne reprend pas exactement les idées de B. Grosz, surtout en ce qui concerne la structure des dialogues, travaille elle aussi de façon descendante pour ce qui est du suivi du thème, bien que cela ne soit pas très clair. En effet, elle aussi construit à chaque phrase un ensemble de thèmes susceptibles d'être développés à partir des termes nouveaux de la phrase et des connaissances contenues dans les variables ("slots") du schéma représentant le thème en cours de discussion, variables qu'elle essaye de remplir.

Par contre, parallèlement à cela, elle écrit qu'un changement de thème ne peut être prévu et qu'il peut être détecté seulement après qu'il a eu lieu (42 p.241).

Mais pour la recherche des référents et co-référents, C. Sidner applique un processus montant, partant de la phrase et des schémas qui lui sont associés pour arriver jusqu'au thème.

Elle n'a pas approfondi non plus un mécanisme de sélection du thème, étant toujours dans un univers d'emploi du temps. Elle doit reconnaître s'il s'agit d'une demande ou d'une assertion et le sujet sur lequel cela porte ("topic").

Pour trouver ce sujet, elle sélectionne le premier groupe nominal suivant le verbe, souvent le complément d'objet; si il n'y en a pas, elle sélectionne le sujet.

Cela est suffisant dans son domaine de travail mais ne constitue pas une règle générale, ce qu'elle reconnaît elle-même.

1.1.3- Conclusion.

En fait, C. Sidner, ainsi que B. Grosz, ne traitent pas les changements complets de sujets. Elles cherchent à reconnaître des changements de thème, mais en restant dans le même thème général. Ceci revient à reconnaître le développement d'un même thème mais sous un autre angle ou bien le développement d'un élément particulier du sujet.

Elles travaillent sur des textes qu'elles supposent cohérents et ne se posent pas le problème de vérifier cette cohérence. Elles l'utilisent au contraire pour délimiter un espace de recherche susceptible de contenir les référents des pronoms et des groupes nominaux définis de la phrase.

Il est certain que ces deux problèmes sont étroitement liés. Cela semble assez circulaire à Hirst (29) selon qui: "Pour déterminer le thème, il faut résoudre les références et pour résoudre les références, on doit connaître le thème". Nous pensons qu'en général, une fois établies les relations qui rendent le texte cohérent, on peut alors trouver les référents. Mais dans

certain cas, pour trouver ces relations, il faut avoir résolu les problèmes de références.

En fait on ne peut établir la tâche qui doit toujours être accomplie la première. Il y a plutôt interaction entre ces modules et cela est propre à chaque cas. Nous exposerons notre position vis à vis de ce problème au paragraphe suivant.

Il faut noter un point particulier. Dans les travaux que nous avons cités, on ne trouve jamais une représentation particulière du thème, de son évolution au cours de l'histoire. Elle se trouve intégrée à la représentation de l'histoire, que ce soit au moyen de réseaux partitionnés ou de schémas.

Ceci signifie que l'on ne prend pas en compte le déroulement d'un sujet pour sélectionner les thèmes possibles, mais que l'on travaille uniquement sur toutes les informations que l'on connaît, qu'elles aient été explicitées dans l'histoire, ou inférées à partir du corpus de connaissances que l'on a répertorié.

Or, la façon d'enchaîner les sujets constitue une aide supplémentaire pour suivre un thème et nous ne négligeons pas cette connaissance dans notre travail.

Comme nous nous intéressons à la vérification de la cohérence, ce sont les problèmes de sélection du thème du discours, de suivi de ce thème, de changements de sujet que nous étudierons dans ce travail.

Nous avons vu que ces problèmes ont été peu abordés

et quand ils l'étaient, on les a grandement simplifiés. Ces simplifications proviennent d'une délimitation stricte des domaines d'études choisis. Nous allons exposer maintenant comment nous traitons la sélection du thème d'un discours alors que nous ne disposons d'aucune connaissance a priori sur les textes eux-mêmes.

1.2- Sélection du thème du discours.

Le problème de la sélection du thème est un point très délicat. En effet, de façon extrême, on ne connaît vraiment le sujet d'un texte qu'une fois sa lecture terminée. Il peut, jusqu'à la dernière ligne, se produire un retournement de situation tel que l'on s'aperçoit que le sujet n'est pas du tout ce que l'on croyait; (exemple de ce phénomène, la chute d'une histoire).

Mais, et c'est ce qui nous intéresse ici, avant d'arriver à la chute, il faut avoir cru en un sujet afin de pouvoir lire l'histoire bien sûr, mais aussi afin de l'apprécier.

On peut étudier la plupart des textes sous les deux points de vue suivants:

On interprète les termes au premier degré, et le texte traite d'un sujet.

On peut aussi considérer le tout au second degré, on reprend alors le premier sujet trouvé, on le comprend sous un autre angle, et le texte traite alors d'un sujet plus général.

Nous nous limitons à la recherche de la première

interprétation d'un discours, celle qui est la plus immédiate. Le sujet exposé constitue le thème principal, dans la mesure où tout le texte est supposé traiter de ce sujet.

Cette limitation de notre système est due à une différence du processus d'analyse, du raisonnement pour trouver les deux types de thèmes. La recherche du thème général demande d'autres connaissances - à propos de l'auteur, de l'époque où il écrit ... - que celles que nous utilisons pour rechercher le thème principal et suivre son évolution. A ce niveau de compréhension d'un texte, il faut mettre en oeuvre un processus de réflexion approfondie sur un sujet.

Nous avons constaté un principe, qui nous semble vérifié par la majorité des textes, et qui permet de trouver le thème principal. Ce principe est le suivant:

Lorsqu'une personne veut parler d'un sujet, elle l'introduit par une phrase et le confirme ensuite*.

Au début du traitement d'un discours, nous considérons que le thème principal est le premier sujet qui est repris dans deux phrases consécutives.

Avant de poursuivre, il nous faut définir le lien qui existe entre une phrase et un thème. En effet, à partir des phrases, donc des termes qui y sont présents, nous allons extraire le sujet qui y est traité.

Nous avons vu que pour C. Sidner, le thème du discours est le groupe nominal suivant le verbe, ou s'il

*-----

Nous ne traitons pas les textes consistant en une énumération de faits ou d'actions n'ayant qu'un rapport épisodique entre eux, ou reliés seulement par des liens de surfaces (du type marabout, bout de ficelle ...). Leur principe de cohérence est différent de celui que nous exposons ici et repose sur d'autres critères.

n'existe pas, le sujet. Pour I. Reinhart, (37), le thème d'une phrase peut être n'importe lequel de ses termes.

Nommer le thème au moyen d'un concept unique, que ce soit le thème du discours ou de la phrase, ne lui donne pas une signification précise, sauf si on travaille dans un domaine restreint où un concept a en fait un sens très particulier.

En outre, cette méthode de sélection d'un thème est trop superficielle. Il est alors équivalent de dire:

P1 : Je vais au restaurant

P2 : J'achète un restaurant

On en extrait le même thème : Restaurant, alors qu'on ne va certainement pas développer le même sujet.

Extraire un seul concept d'une phrase, quelle que soit sa fonction sémantique, ne permet pas de nommer le sujet très précisément. Nous considérons que le prédicat joue de toute façon un rôle très important et qu'il faut introduire cette information dans le nom du thème. C'est pourquoi nous extrayons deux éléments de la phrase: le PREDICAT et son OBJET.

Le sens de ce couple est le suivant:

Le prédicat indique le point de vue sous lequel on considère l'objet.

Le choix de l'objet dans la phrase dépend du prédicat et des cas sémantiques qui le définissent.

Par exemple, pour des verbes d'action transitifs, ce sera le cas sémantique "thème" (souvent complément d'objet direct); pour des verbes de déplacement, ce sera le cas "lieu" ...

Dans notre exemple, le thème de P1 est "aller au restaurant" et celui de P2 "acheter un restaurant". Nous indiquons ainsi immédiatement comment le sujet de "restaurant" est abordé.

C'est à ce point de l'analyse que peut se poser le problème de connaître les référents des groupes nominaux et des pronoms.

1.3- Cohérence et références.

Le problème de la recherche des référents est différent de la vérification de la cohérence, car ce n'est pas l'emploi de pronoms ou de groupes nominaux définis qui provoque la cohérence d'un texte, si ce n'est une cohérence de surface. Ils en découlent plutôt.

Selon Hobbs (30), la stratégie qu'adopte un auditeur pour comprendre un discours, consiste à faire de son mieux pour reconnaître la cohérence, puis de faire des suppositions de co-références qui gardent le principe de cohérence trouvé. Les problèmes de références se résolvent alors par identification de variables.

Nous avons adopté une stratégie similaire à celle-ci. Trouver le raisonnement qui relie une phrase à un contexte connu permet de préciser le champ de recherche

des référents. Ce raisonnement peut être trouvé à l'aide des concepts génériques. Il n'y a donc pas de problème en ce qui concerne les groupes nominaux.

Les pronoms, lorsqu'ils remplacent l'objet dans la phrase, sont plus délicats à traiter car nous ne pouvons pas extraire de la phrase le couple significatif. Nous devons alors rechercher l'antécédent du pronom^{*}; cela peut se faire à partir de critères syntaxiques et sémantiques, Sabah (39).

Les critères syntaxiques sont du type genre et nombre du pronom, ou la place des antécédents possibles dans la phrase.

Les critères sémantiques proviennent des cas attachés au prédicat et des traits que possèdent les éventuels antécédents. Ceux-ci permettent de résoudre les cas de figure suivants:

Exemples tirés de la thèse de G. Sabah.

"Le chien qui voit un sucre, monte sur le buffet. Il veut le manger."

"Dans le jardin, un chat voit une souris. Il se glisse derrière une laitue car il veut la manger."

Il se peut que certains pronoms ne possèdent pas d'antécédents ou qu'ils n'aient pu être compris. Nous suggérons alors de procéder de la manière suivante. Nous remplaçons ce pronom par un objet par défaut lié au prédicat et la recherche de la cohérence peut avoir lieu. Nous nous ramenons ainsi au cas où la phrase ne

* Nous rappelons la différence entre le référent d'un pronom, qui est l'objet du monde réel auquel ce pronom fait référence, et l'antécédent de ce pronom, qui constitue une spécification de l'objet, précédemment mentionnée dans le discours.

Le processus décrit est simulé dans notre programme.

mentionne aucun objet.

Nous donnons ainsi un sens plus général à la phrase, sens qui sera précisé quand le référent aura été trouvé par une recherche analogue à celle appliquée aux groupes nominaux.

Dans le dernier cas exposé, l'ordre des tâches est inversé. Il est préférable de rechercher les antécédents d'abord, puis de vérifier la cohérence et trouver le thème traité.

De ce fait, la résolution des pronoms ne peut pas reposer sur un principe de cohérence du texte et c'est pourquoi nous utilisons d'autres moyens, fondés sur d'autres critères pour y parvenir. Et seulement dans le cas où ce processus n'aboutirait pas, ou que le pronom ne posséderait pas d'antécédent, il faudrait rechercher le référent en fonction du thème développé, en supposant qu'on fait référence à un objet normalement en mémoire.

Examinons maintenant comment se déroule le suivi d'un thème. La connaissance du thème principal va nous permettre de situer par rapport à lui les différents sujets développés dans le texte.

1.4- Déroulement du thème.

La présence d'un thème principal ne supprime pas la possibilité de parler d'autres sujets. Nous avons tout à fait conscience que nous pouvons trouver des digressions, des parenthèses, c'est à dire des changements de thème, ainsi que le développement d'un point particulier

d'un sujet.

Chaque écart d'un thème peut être définitif - on ne revient jamais au thème qui était développé avant le changement - ou au contraire momentané.

Rien ne peut nous indiquer quand se produisent ces phénomènes.

Contrairement à B. Grosz qui peut reconnaître quand les différents sujets sont clos, puisqu'il lui suffit de reconnaître qu'une tâche est terminée, nous n'avons pas pu introduire ce type de processus, n'ayant pas restreint notre étude aux seuls textes offrant ces possibilités.

C. Sidner, quant à elle, considère qu'un sujet est clos lorsque toutes les caractéristiques en sont données; cela signifie qu'alors le schéma représentant le thème est rempli.

Elles considèrent toutes deux que lorsqu'un sujet est clos, on ne peut plus y revenir par la suite.

Nous n'avons pas introduit de telles règles. Nous considérons que l'on peut reparler d'un sujet même si celui-ci semblait épuisé. Par contre, nous considérons que revenir à certains sujets, alors que ceux-ci ont été abandonnés et qu'ils sont totalement différents du thème principal ou de celui qui est en cours de développement constitue une incohérence; en effet, la plupart des personnes lisant un texte de ce type seraient interloquées et ne comprendraient pas.

Si le discours fait explicitement mention d'un retour à un ancien sujet, ceci sera alors accepté.

Nous considérons par contre, que l'on peut revenir au thème principal à tout moment du discours.

Ces principes sur le développement d'un sujet nous permettent de sélectionner, à chaque phrase du texte, un ensemble de thèmes auxquels devra se rapporter la phrase étudiée afin de respecter les principes de cohérence.

C'est ainsi que nous décrivons le contexte dans lequel nous cherchons à intégrer la phrase analysée. Nous reviendrons à la description du contexte lorsque nous aurons précisé ce que signifie "intégrer une phrase dans un contexte".

Nous avons parlé de digression, de parenthèse, de changement de thème. Nous allons définir plus précisément les différents procédés permettant de développer un sujet en restant cohérent avec le reste du texte.

Dans les travaux déjà cités, nous avons vu que les changements de thème sont limités au développement d'un élément particulier du sujet.

Dans ce travail, nous nous situons dans un cadre plus général. Aussi nous avons recherché les différents types de changement de thème admissibles pouvant se produire dans un texte.

Nous avons établi trois catégories de processus. Etant donné un sujet en cours de traitement, il est possible de:

- continuer à parler du sujet en cours;

Ex: Je suis allé au restaurant. J'ai pris le menu à 150 francs.

- développer un point particulier d'un sujet: on parle toujours du même thème, mais sous un point de vue différent. Nous appelons cela une déviation de thème;

Ex: Je suis allé au restaurant. Il est "style année 30". Il y a des miroirs partout.

- parler d'un autre sujet, mais celui-ci possède tout de même un point commun avec le sujet traité. Nous appelons ce dernier processus un changement de thème, car il y a introduction d'un nouveau sujet.

Ex: Je suis allé au restaurant. J'ai rencontré Marcel. Il a acheté une voiture.

Nous avons vu que nous décrivons un contexte avant de traiter chaque phrase. L'intégration de la phrase dans un contexte signifie la reconnaissance de l'un de ces enchaînements par rapport aux thèmes faisant partie du contexte.

Si un tel lien existe, il permet de situer la phrase par rapport au reste du texte, de décrire le raisonnement rendant le tout cohérent.

Supposer que ce lien puisse ne pas exister, donc

que le texte soit incohérent, nous amène à en faire la recherche de façon montante : nous partons de la phrase pour tenter d'arriver à ce qui a déjà été analysé. Nous verrons plus loin que le processus d'analyse n'est pas totalement montant, mais qu'il est en partie descendant.

Ce principe d'une analyse montante est très important, autant par sa signification vis à vis de la cohérence d'un texte que pour la démarche, le type de raisonnement qu'il traduit. Nous développerons ce point lorsque nous aurons examiné tout le processus d'analyse.

Voici un exemple permettant d'illustrer le suivi d'un thème.

M. est garagiste.

C. - { Monsieur, j'ai un canot automobile dans le port et
<1> { mon moteur est en panne. Une espèce de pêcheur-
matelot-joueur-de-boules sur le quai, m'a dit de
m'adresser à vous.

M. - C'est bien gentil de sa part. Mais je ne sais pas
si je peux y faire grand chose.
<2> { Je m'occupe surtout de voitures, n'est-ce-pas. Je
n'ai pas fait d'études spéciales.
Alors si c'est un Diesel.

C. - Ce n'est pas un Diesel. C'est un Beaudoin.

M. - Ah! les Beaudoin, je les connais un peu. J'en ai soigné un pendant trois ans.

- Il était sur le bateau de M. Frère, le Président
du Tribunal. Vous le connaissez ?
- <3> {
C. - Non, je ne suis pas d'ici.
M. - Oui, enfin ça ne fait rien. On va voir votre outil.
Ficelle, si on me demande, je suis au port sur le ...
{ Comment il s'appelle votre bateau ?
<4> {
C. - Le Pescadou.
M. - Sur le Pescadou. C'est un joli nom.
Attendez, je vais prendre un jeu de clefs.

Exemple I.2: Extrait de César de M. Pagnol.

- <1> Permet de faire une hypothèse sur le thème de ce discours: on va parler d'un moteur de bateau en panne et de sa réparation par un garagiste.
- <2> Déviation du thème principal dans la mesure où le garagiste parle plus particulièrement de lui.
- <3> Changement de thème complet. On y parle d'une tierce personne n'ayant pas de rapport avec le thème du discours, mais on y arrive de façon cohérente et on peut trouver un lien avec le sujet principal.
- <4> Nouvelle déviation de thème, en parlant du nom du bateau.

Remarquons qu'il est fait mention d'un "outil". Nous verrons comment le thème intervient dans le traitement de ce mot (§III.6).

La connaissance des emboîtements des différents thèmes les uns par rapport aux autres influence la description du contexte. Si l'on est à l'intérieur d'une parenthèse - c'est-à-dire qu'un changement de thème a eu lieu - le contexte n'est pas décrit comme lorsqu'il y a déviation de thème. La différence réside dans les sujets que nous considérons cohérents de retrouver, de près ou de loin.

I.5- Le contexte.

Nous avons vu que le thème peut varier au cours de l'analyse; aussi nous possédons la notion de thème actif, "thème immédiat" chez B. Grosz; c'est celui qui est en cours de développement.

Le contexte est donc constitué d'un ensemble de thèmes; il contient systématiquement le thème principal et le thème actif, s'ils sont différents.

De plus, lorsque l'on est en cours de changement de thème, on y ajoute le thème qui était actif au moment où ce changement s'est produit. Ce qui signifie qu'il est cohérent de revenir au sujet traité avant la parenthèse, même sans le mentionner explicitement.

Reprenons l'exemple I.2: Le thème principal en est "réparer un moteur".

Au niveau <2>, le thème actif est "être garagiste". Le contexte est alors constitué du thème principal et du thème actif.

Au niveau <3>, le thème actif est "connaître quelqu'un". Il y a eu changement de thème. Le thème actif avant ce changement était "soigner un moteur". Ici le contexte est formé de ces deux thèmes et du thème principal.

Se définir ainsi un contexte, c'est émettre des hypothèses sur le thème auquel se réfère la phrase. Ceci constitue la partie descendante dans notre analyse du texte.

Tous ces processus nous permettent de reconstituer la structure du texte en trouvant les enchaînements, les liens entre chaque phrase.

Nous pouvons alors connaître les inférences à faire afin de comprendre les phrases.

Ces hypothèses ne délimitent pas un cadre étroit dans lequel doit se loger la phrase. Nous avons vu les différents liens possibles; ils visent à respecter le style, les enchaînements de la majorité des auteurs dans la mesure où ils entrent dans notre définition de la cohérence.

Si ces liens ne peuvent pas être trouvés, nous sommes dans un cas d'incohérence, d'incompréhension de la démarche de l'auteur.

Ce principe d'une analyse à la fois montante et

descendante pour mettre un raisonnement en évidence est très important.

I.6- Analyse montante et descendante.

Par la construction d'un contexte avant l'analyse de chaque phrase, nous précisons les limites de la cohérence et surtout celles de l'incohérence.

Nous n'avons pas choisi de rechercher un contexte propre à chaque phrase, mais d'intégrer chaque phrase dans un contexte existant.

Il est à noter que nous ne pré-supposons pas les thèmes à venir, mais que nous supposons cohérent le développement d'un ensemble de thèmes choisis en fonction de ce qui a déjà été analysé. Cet ensemble constitue le point d'arrivée d'une analyse montante des phrases.

Plutôt que partir du contexte et faire toutes les hypothèses possibles afin d'arriver à confirmer l'une d'elles, nous avons choisi de partir de la phrase.

D'un point de vue pratique, cette méthode permet de limiter les inférences. Il est également souhaitable de les limiter d'un point de vue théorique.

Faire de nombreuses hypothèses signifie, comme nous l'avons déjà dit, connaître tous les enchaînements a priori. Cela implique aussi de rechercher et de retenir une grande quantité d'informations jusqu'à ce que l'une d'elles soit confirmée.

Ce principe traduit une démarche très lourde et, à ce qu'il nous semble, assez irréaliste. C'est pourquoi nous avons préféré mettre en place une analyse montante.

Nous pouvons ainsi ne sélectionner que l'information utile à la constitution des raisonnements qui sont indispensables à la compréhension d'un texte; cette sélection a lieu au fur et à mesure des besoins.

Outre la possibilité de choisir uniquement l'information pertinente, nous désirons organiser celle-ci en niveaux de compréhension.

En effet, il y a différentes façon de parler d'un sujet en fonction des informations que l'on veut communiquer. Cela se traduit par le choix de termes propres à traduire une pensée.

Ainsi, lorsque l'on parle de réparer une voiture, on ne veut pas transmettre, par ces termes, le fait de dévisser le troisième écrou à gauche, situé sur la boîte de vitesse, mais plutôt le fait qu'il y a un problème en ce qui concerne le fonctionnement de la voiture, qu'elle est immobilisée pour un temps ...

Si l'on veut préciser un des éléments, il faudra alors le dire explicitement; on descendra alors d'un niveau, ou plus, dans la compréhension du sujet grâce aux détails qui seront donnés.

Nous ne désirons pas tout ramener à une même échelle car nous voulons distinguer, parmi les connaissances,

les détails des généralités.

Nous traduisons cela autant par le processus d'analyse que par notre représentation des connaissances (§ 11.3).

Les processus liés au thème du discours étant précisés, nous allons maintenant étudier plus précisément comment se déroule l'analyse d'un texte.

Nous tenons à rappeler que cette analyse ne vise pas une compréhension totale du texte, des personnages...

Notre but est, pour comprendre chaque phrase, de mettre en évidence les inférences, les raisonnements nécessaires à cette compréhension. Le thème du discours est donc le fil conducteur de cette analyse.

Nous avons vu comment nous sélectionnons le thème principal. Ceci constitue la première étape du processus.

Ensuite nous étudions le texte phrase par phrase et nous cherchons le raisonnement qui les relie à ce qui précède. Pour cela, nous procédons à une analyse syntaxico-sémantique de la phrase, Rady (36), et nous en extrayons les éléments qui nous intéressent: le couple PREDICAT + OBJET, si ce couple existe, sinon l'un des deux concepts.

Ce couple représente la signification de la phrase pendant la recherche du lien.

Le but de cette analyse est aussi de préparer une compréhension plus approfondie du texte. Aussi le problème de la représentation du discours s'est posé sous ce point de vue et non pas sous celui développé par Kintsch et Van Dijk (31) qui représentent les textes au moyen de macro-structures.

Nous ne cherchons pas à établir une représentation du texte telle qu'elle peut se trouver en mémoire.

Nous reproduisons seulement la signification du texte sous le point de vue de son thème.

Cette représentation devant être le point de départ à une analyse d'un autre ordre, nous avons cherché à organiser les informations tirées des textes.

I.7- Représentation du discours.

Les informations à représenter sont de deux types. Premièrement celles qui découlent de l'étude du déroulement du thème, ensuite celles qui proviennent du contenu des enchaînements mis en évidence.

Les premières donnent des informations sur la structure du texte. Nous conservons celles-ci afin de connaître, à chaque pas de l'analyse, les thèmes qui ont été développés et surtout ceux qui sont en cours de développement afin de constituer le contexte.

Il faut connaître les types d'enchaînement entre les sujets: s'il s'agit d'une confirmation, d'une

déviations, d'une inférence ou d'un changement de thème.

Pour cela, nous représentons le déroulement du discours sous forme d'arborescence. Un noeud est un thème explicitement abordé dans le discours, ou bien un thème implicite, c'est-à-dire une inférence. Les liens entre chaque noeud représentent le type d'enchaînement entre deux sujets.

Les informations du second type nous indiquent comment les sujets sont développés, à savoir si le thème est introduit de façon générale pour être précisé ensuite; ou inversement, si l'on a d'abord exposé des points particuliers du sujet.

Ces enchaînements reflètent également les relations de cause à effet ou l'inverse. Une autre manière d'enchaîner deux phrases est de donner des explications d'un fait.

La connaissance de la nature des développements des sujets est bien sûr indispensable à la compréhension des textes. Nous verrons comment nous reconnaissons ces liens et comment nous les représentons lorsque nous décrirons la représentation des connaissances que nous avons mise en place: cette représentation est fondée sur la notion de schéma.

Nous allons étudier maintenant quelques cas que la connaissance du thème du discours permet de traiter. Nous n'indiquerons pas dans ce chapitre comment nous résolvons le problème. Nous en donnons une définition et nous évoquons des solutions possibles.

I.8- Ambiguïtés étudiées par le biais du thème.

Lorsque nous extrayons d'une phrase le couple prédicat-objet, nous partons d'une phrase où l'analyse syntaxico-sémantique est faite. Nous connaissons donc les concepts que recouvrent les mots employés. Nous développerons ceci au § II.1.

Signalons seulement qu'à un mot peut correspondre plusieurs concepts, plusieurs sens, et qu'à plusieurs mots peuvent correspondre un même concept dans les cas de synonymie.

Il est des cas où la phrase est ambiguë dans le sens où les seules connaissances syntaxiques et sémantiques ne permettent pas de choisir le sens de l'un au moins des mots la composent.

Ex. extrait de l'exemple I.2:

Je vais prendre un jeu de clefs.

Les mots "jeu" et "clefs" ont plusieurs sens parmi lesquels il est impossible de faire un choix.

C'est alors que la connaissance du thème intervient. Si la phrase est ambiguë à ce niveau de l'analyse, la connaissance du contexte dans lequel elle arrive, le thème actif avant son arrivée peuvent nous permettre de la comprendre.

A partir des concepts connus dans le contexte, on

pourra rechercher le sens adéquat du mot ambigu. Si une recherche de ce type n'aboutit pas, l'ambiguïté reste.

Un autre cas d'ambiguïté se produit quand l'analyste trouve que des contraintes sémantiques, les cas liés au prédicat, sont en contradiction avec les concepts de la phrase.

Ex. extrait de l'exemple I.2:

J'en ai soigné un (un moteur) pendant trois ans.

Nous connaissons deux sens à "soigner":

- soigner un être animé
- soigner un travail

Un prédicat est en effet défini sémantiquement par une liste de cas représentant les différents compléments que l'on peut trouver avec ce concept. Ces cas permettent de décrire les contraintes sémantiques que doivent remplir ses compléments (voir § II.1).

Quelle est la signification d'une telle situation ?

Nous savons que la phrase représente une idée de son auteur, qu'elle a un sens pour lui. Si nous ne pouvons trouver ce sens, cela provient sans doute de notre définition du prédicat.

On peut expliquer ainsi ce phénomène:

- Soit le prédicat est employé ainsi parce qu'il n'existe pas d'autre concept pour exprimer une idée et alors

nous sommes en face d'un sens du prédicat inconnu du système.

- La seconde possibilité vient de ce que le locuteur emploie le prédicat en sur-extension, Anglin (1).

Cette utilisation, correspondant à un emploi métaphorique du prédicat, signifie que l'auteur applique ce prédicat à un objet qui possède des caractéristiques autres que celles qui en définissent l'emploi usuel; c'est-à-dire qu'il l'utilise avec un éventail d'objets plus vaste que celui connu du système.

Dans ce cas, le prédicat a un sens que le système connaît sans doute, mais qu'il désigne sous un autre nom, par un autre prédicat.

L'emploi d'un concept à la place d'un autre ne nous semble pas dû au hasard mais à une synonymie partielle de ces deux concepts dans un contexte déterminé.

Il nous faudra donc retrouver le prédicat que notre système connaît et qui peut s'adapter à une telle situation.

Nous verrons comment nous résolvons ce problème au § III.5. Nous pouvons préciser que la solution est due à la représentation du thème du discours sous forme de schémas, à la connaissance du contexte et du thème, qui permettent de choisir un sens possible pour le prédicat.

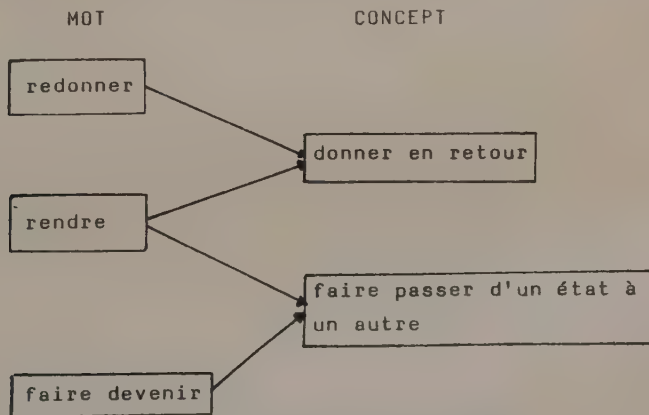
Chapitre II

Représentation des connaissances

II.1- Représentation des concepts

Nous nous intéressons ici à la représentation du sens des mots par des concepts. La représentation des informations morphologiques et syntaxiques des mots ne présentent aucune utilité pour notre travail.

Nous tenons à souligner la différence entre mots et concepts. Un mot peut renvoyer à plusieurs concepts (cas de polysémie). D'autre part plusieurs mots peuvent correspondre à un même concept (cas de synonymie).

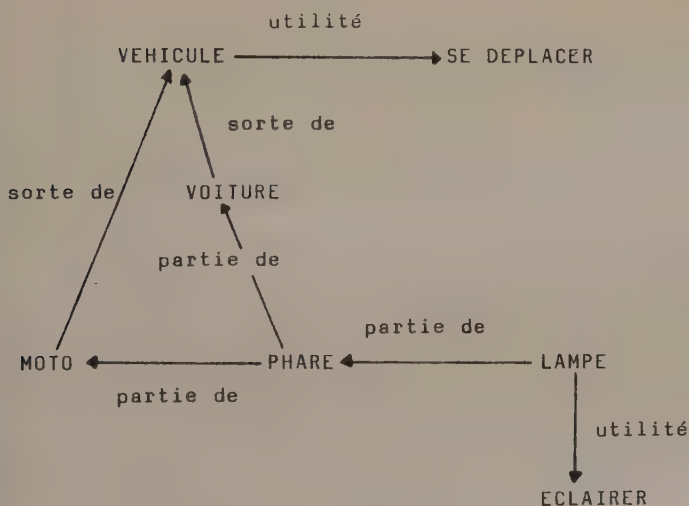


Exemple II.1- Relation mots-concepts

Nous avons choisi de représenter les concepts au moyen d'un réseau sémantique Simmons (45), Fahlman (19). Les relations entre chaque noeud -chaque concept - sont de deux types:

- un lien hiérarchique (sorte de); il permet de transmettre les propriétés par héritage;
- les qualités attachées à un concept (partie de ...).

Certaines de ces relations sont transitives.



Exemple II.2- Extrait du réseau sémantique

Les concepts prédicatifs sont définis par des restrictions portant sur leurs arguments qui sont répertoriés par six cas: source, agent, thème, instrument, lieu et but Fillmore (21).

soigner

agent	: humain
thème	: maladie
but	: animé
instrument	: médicament, traitement

Exemple II.3- Représentation d'un prédicat

Nous avons besoin de faire cohabiter un réseau sémantique avec des schémas pour deux raisons:

- lorsqu'un concept ne constitue pas une entrée vers un schéma, il faut rechercher un concept plus général qui permette de trouver un couple formant le nom d'un schéma. En effet, il y a plusieurs concepts synonymes pour une même situation.

schéma : réparer véhicule

Ce schéma est le même pour :

- * réparer voiture
- * réparer moto
- * réparer camion

Si l'on trouve l'un de ces concepts dans une phrase, il faut alors le généraliser. C'est une opération très simple, fondée sur les relations hiérarchiques du réseau.

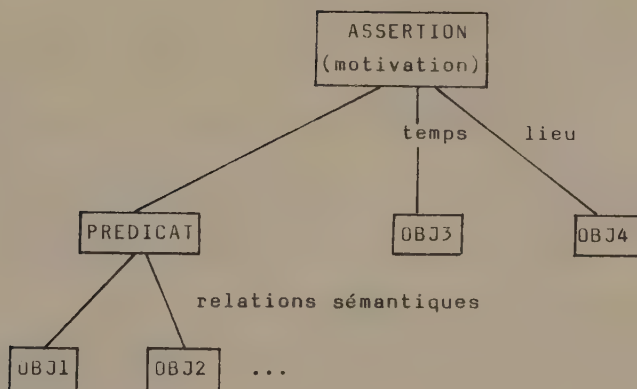
- lorsqu'un concept fait partie d'un schéma, nous le décrivons par certaines de ses qualités. Ce sont les qualités précisées par le contexte.

II.2- Représentation interne des phrases

Nous avons déjà mentionné que nous travaillons à partir des phrases analysées. Nous avons donc choisi une représentation interne pour ces phrases. Celle-ci est analogue à ce qui existe dans la thèse de G. Sabah.

Notre but n'étant pas de statuer sur la validité psychologique d'une telle représentation, nous en avons choisi une permettant de servir de point de départ à notre système et pouvant être créée par l'analyseur de M. Rady.

Ainsi que nous l'avons précisé § 1.3, nous partons d'une représentation où les concepts sont identifiés. S'il y avait pronom dans la structure de surface, il a été remplacé par le concept qu'il reprend. Lorsque ces problèmes n'ont pu être résolus, nous nous trouvons dans l'un des cas particuliers décrits § 1.8.



Exemple II.4- Représentation interne d'une phrase

La motivation permet de distinguer le type de la phrase. Celle-ci est soit une information (motivation affirmative, négative, restrictive), soit une demande ou une condition.

Le prédicat contient le qualificatif de la phrase; il est relié à des objets par les différents cas sémantiques. Ces objets sont des occurrences de concepts dont on possède une représentation dans le réseau et dont certains traits peuvent être explicités dans la phrase.

Chaque objet peut éventuellement être précisé par une ou plusieurs autres phrases. La motivation indique alors la nature du lien entre les deux phrases.

II.3- Les schémas

II.3.1- Description d'un schéma

Nous allons étudier différents points concernant les schémas, à savoir ce qu'ils représentent, pourquoi utiliser des schémas et ce qu'ils signifient.

Notre travail repose sur la notion de thème du discours. Nous l'avons défini comme un sujet de discussion qu'il est possible de développer de différentes façons: on ne parle que de ce sujet, on en développe un point particulier ... Il nous faut donc regrouper tout ce qui fait partie d'un thème.

1) Connaissances représentées dans un schéma.

On peut se représenter un sujet comme une situation relatant des événements ou évoquant des notions abstraites. C'est cette association qui nous a conduit à choisir des schémas, ceux-ci permettant de représenter de manière assez complète les multiples relations entre les éléments d'un thème et les liaisons entre plusieurs thèmes.

Nous avons construit une base de connaissances regroupant un certain nombre de schémas qui constituent des stéréotypes. Ces schémas contiennent les informations suivantes:

- le nom du schéma;
- la description de la situation: la suite des actions à effectuer ou la description d'un état;

- les conditions pour que la situation se produise;
- les résultats après que la situation a eu lieu;
- le lieu où se déroule la situation, si ce lieu est une donnée propre à cette situation;
- le temps, il peut être une date, une fréquence, si cette information est aussi une donnée propre à la situation.

Il est évident que la présence ou l'absence de chaque type d'informations dépend de la situation à décrire.

Quels détails allons nous donner de la situation et quelles sont les connaissances que nous voulons y trouver ?

Nous avons vu § 1.6 que nous désirons conserver les informations transmises par leur auteur en niveaux de compréhension pour garder une trace de la démarche de cet auteur.

Nous voulons aussi sélectionner uniquement l'information utile à notre compréhension des phrases.

Les schémas vont nous permettre de réaliser ces processus. Contrairement à leur définition usuelle qui consiste à regrouper dans une même structure tout ce que l'on sait d'un sujet dans ses moindres détails ou tout ce qui vient à l'esprit lorsque l'on évoque un thème, nous avons essayé de faire éclater une telle structure.

Les critères que nous avons choisis pour répartir ces connaissances nous ont été inspirés par notre analy-

se du texte. Un schéma représente donc un degré dans la compréhension d'une situation et constitue un élément de la décomposition d'un sujet en plusieurs étapes, allant du général au particulier.

Outre la signification que cette représentation a pour l'analyse du texte, cela nous permet de limiter la masse de connaissances regroupées sous un même nom. Nous évitons ainsi une description trop rigide des situations telles qu'ont pu en construire Charniak (14) et Schank (40) au moyen des "frames" et des "scripts".

Nous venons de parler du nom d'un schéma. C'est en effet une information que nous possédons pour chaque schéma; il va nous permettre d'accéder au schéma. Voici un nouvel avantage à la décomposition d'une situation en plusieurs niveaux: on peut leur donner un nom représentatif. Ce nom est constitué de deux éléments: un prédicat et son objet.

Nous avons déjà mentionné un tel couple au § 1.2. C'est celui que nous extrayons d'une phrase pour en résumer la signification et chercher à la relier à la partie précédente du texte. Il est donc normal de retrouver ce couple quand il s'agit de nommer une situation, et c'est toujours par son nom que nous atteindrons un schéma, que nous le déclencherons* à partir des phrases.

Nous allons préciser maintenant la forme d'un schéma.

* la signification de ce terme sera donnée au chapitre suivant.

2) Structure d'un schéma

Nous avons vu que chaque schéma peut contenir les parties description, conditions, résultats, lieu et temps en plus du nom. Chacune de ces parties correspond à une structure regroupant plusieurs éléments. Chaque structure possède un nom qui la précise et indique son rôle par rapport aux autres structures dans le schéma. Ce nom est celui des parties énumérées ci-dessus.

Les éléments d'une structure représentent les valeurs qui lui sont attribuées. Elles sont de deux natures:

- un autre schéma (en fait un pointeur vers un schéma)
- une variable: c'est la référence à un concept du réseau sémantique ou à un sous-ensemble de ce réseau (un concept principal et la sélection de certains traits et descendants).

Ainsi, lorsque la valeur d'un élément est trop complexe à décrire, nous faisons référence à un schéma. Il représente alors un des points de la situation développé par ailleurs.

A chaque élément, sont associées des procédures. Elles ont des fonctions diverses. Certaines traduisent les conditions que les concepts doivent satisfaire pour qu'ils aient un sens dans le contexte décrit. D'autres précisent les concepts "transportables" quand la valeur est un schéma, et leur fonction, leur signification dans le schéma appelé.

NOM : réparer véhicule (cas thème)

AGENT : VALDEF: personne (obl)

INSTRUMENT : VALDEF: outil (fac)
 TRAIT(outil,nombre,x)
 TRAIT(outil,possesseur,agent)

CONDITION : N°DIST N°ORDRE
 1 0 * véhicule en panne (obl)
 EGAL(véhicule,compl)
 2 0 * connaître mécanique (obl)
 EGAL(agent,agent)

DESCRIPTION : N°DIST N°ORDRE
 1 1 * trouver panne (obl)
 EGAL(agent,agent)
 2 2 * démonter pièce (fac)
 EGAL(pièce,l.res)
 EGAL(agent,agent)
 EGAL(lieu,lieu)
 3 3 * réparer pièce (obl)
 EGAL(pièce,l.res)
 EGAL(agent,agent)
 4 4 * remonter pièce (fac)
 EGAL(pièce,l.res)
 EGAL(agent,agent)
 SIFAIT(1,obl)

RESULTAT : N°DIST N°ORDRE
 1 0 * véhicule fonctionne (obl)
 EGAL(véhicule,compl)

LIEU : sorte-de local (fac)
 VALDEF: garage

Exemple II.5- Schéma "réparer un véhicule"

Le nom du schéma est représenté par deux concepts et le nom du cas qui relie l'objet au prédicat.

A cela s'ajoutent les autres cas du prédicat qui ne sont répartis ni dans les structures du schéma comme le lieu, ni dans son nom. Ils sont représentés par des structures dont le nom est celui du cas et la valeur en est le concept de la phrase (ex. INSTRUMENT).

Les éléments de chaque structure sont numérotés de deux façons:

- Un numéro (n°dist) permet de les distinguer et ainsi de les nommer à l'intérieur du schéma (cf élément 2 de la structure DESCRIPTION).
- Le second numéro (n°ordre) leur donne un ordre chronologique si cela est nécessaire.

Remarquons que nous possédons ainsi dans un schéma deux types d'enchaînement des événements: chronologique, à l'intérieur d'une structure et causal, par la nature des structures.

Les procédures dont nous faisons mention plus haut sont "TRAIT" et "EGAL" dans l'exemple. "TRAIT" permet de préciser le sens du concept "outil". "EGAL" donne des valeurs aux cas du prédicat, mais aussi aux structures des schémas appelés. Elles sont attachées aux différents éléments du schéma.

La procédure "SIFAIT", intervient uniquement dans les éléments facultatifs. Elle est déclenchée lorsque l'élément est rempli. Elle peut forcer la valeur d'un

autre élément, lui aussi facultatif.

Nous parlerons plus loin d'autres procédures, intervenant au moment du remplissage du schéma.

Les éléments possèdent une information supplémentaire, concernant le caractère obligatoire ou facultatif de leur présence. On peut trouver, dans la structure "CONDITION", des éléments facultatifs. Ceci signifie alors que s'ils sont absents, ou non vérifiés, la situation décrite a toujours un sens et peut tout de même avoir lieu normalement.

Dans les autres structures, l'absence éventuelle des éléments facultatifs ou leur négation ne peuvent pas être considérés comme un déroulement anormal de la situation.

Les éléments obligatoires doivent toujours être présents sinon la situation décrite ne correspond pas exactement à celle que l'on connaît. Nous n'avons pas développé la signification d'un tel cas, puisque nous ne cherchons pas à deviner le pourquoi de ce qui se passe. Nous voulons uniquement mettre en évidence les liens entre les événements.

Lorsqu'un élément est nié, cela prouve quand même un rapport avec le schéma et cela est suffisant pour notre travail.

Ex: Je m'occupe surtout de voiture, n'est-ce pas.
Je n'ai pas fait d'études spéciales.

Nous cherchons uniquement à trouver le rapport entre "faire des études" et "être garagiste", le thème actif à ce moment du discours.

Une autre caractéristique apparaît dans les éléments du type variable: la valeur par défaut, "VALDEF". Cette valeur par défaut est attribuée à l'élément si aucune autre valeur n'a été donnée à cet élément et si elle est nécessaire à la compréhension du discours.

Nous allons décrire maintenant les opérations indispensables à l'utilisation des schémas: l'appel d'un schéma, son déclenchement et son remplissage.

3) Opérations liées à un schéma

. Appel d'un schéma

L'appel d'un schéma, c'est la recherche de celui-ci à partir des phrases, c'est la façon dont nous l'identifions. Nous avons déjà dit qu'un schéma est trouvé grâce à son nom lorsque celui-ci apparaît dans une phrase. Ce n'est pas toujours le nom exact du schéma qui y figure. Il est souvent nécessaire de généraliser le concept objet par l'intermédiaire du réseau sémantique pour trouver le nom d'un schéma. Un tel appel se produit si l'on parle explicitement d'un sujet: appel explicite.

Nous avons vu qu'il est possible de parler implicitement d'un thème, sans le nommer. Dans ce cas nous allons trouver le schéma modèle correspondant par

l'intermédiaire des relations entre deux thèmes explicites: appel implicite. Le schéma ne correspond à aucune réalité décrite par des mots. Il représente une étape dans un raisonnement. Le nom ne joue pas un rôle prépondérant dans ce type d'appel et cela entraîne deux manières de déclencher un schéma.

. Déclenchement d'un schéma

Déclencher un schéma signifie créer une occurrence d'un schéma modèle*. Un schéma instancié possède une structure générale identique à celle du schéma modèle. On peut y préciser en plus le lien qui le relie à une autre occurrence, ainsi que la nature de ce lien, si ce lien ne fait pas partie de ceux qui sont prévus. Nous verrons l'utilité de ces éléments au paragraphe suivant.

Lorsque l'instanciation a lieu à partir d'un appel explicite, les problèmes de références ont été résolus. Ainsi le nom de l'occurrence du schéma modèle est composé des concepts instanciés de la phrase. C'est le seul constituant du schéma que nous précisons au moment du déclenchement. Nous ne recopions aucune des structures, ni aucun des éléments du modèle.

Lorsque l'instanciation a lieu à partir d'un appel implicite, nous ne donnons pas de valeur au nom, faute de le connaître réellement.

Dans ce type d'appel, nous l'avons vu, le schéma exprime une étape d'un raisonnement et ne décrit pas une situation réelle. Aussi le nom ne correspond pas à des

* Nous appelons cette opération "instancier". De même une instanciation est la création d'une occurrence. Nous appliquons aussi ces termes aux opérations équivalentes sur les concepts.

objets réels. Nous ne recopions évidemment pas non plus les composants du schéma modèle. Les schémas ainsi déclenchés pourront être remplis par la suite, s'il leur est fait explicitement référence ou si les schémas remplissent des éléments d'autres schémas. Il se peut alors que des procédures précisent le nom, ou d'autres éléments du schéma inféré.

. Remplissage d'un schéma

Les occurrences sont remplies, complétées par les éléments de la phrase à l'origine de leur appel et par le déclenchement de schémas qui les composent.

Les éléments extraits d'une phrase sont de deux natures.

Il peut s'agir de variables; ce sont alors les "OBJETS" que nous avons décrits dans la phrase. Nous cherchons les éléments du schéma pouvant être remplis par ces variables. Le nom du cas reliant un objet au prédicat ou à la phrase nous guide dans cette recherche et permet de trouver la structure décrite. Nous devons alors faire coïncider la variable de la phrase avec un concept appartenant à l'un des éléments de la structure trouvée.

Cette correspondance se déroule en deux étapes.

Premièrement, trouver un concept du schéma équivalent, du point de vue sémantique, à celui de la phrase; les deux concepts génériques sont les mêmes ou l'un est plus général que l'autre. Nous acceptons cette relation

dans les deux sens: la variable la plus générale est celle de la phrase ou celle du schéma.

Nous vérifions ensuite si les éventuelles contraintes imposées à l'élément du schéma sont respectées par le concept de la phrase. Si ce n'est pas le cas, nous cherchons un autre élément. Si aucun élément n'est trouvé, nous sommes dans l'incapacité de résoudre le problème à l'aide des seules connaissances que nous possédons.

Des situations de ce type sont actuellement étudiées par J.B. Berthelin (3). Il a ainsi établi qu'il faut recenser, pour une application, une situation, les connaissances conceptuelles et les connaissances épisodiques. Ces dernières constituent les points contestables, c'est-à-dire les points où on admet tout à fait des avis contradictoires. On pourra ainsi trouver l'origine de la contestation, sa signification et agir en fonction de cela.

Dans le cas présent le fait de ne pas trouver où placer un concept alors qu'il devrait se rattacher au thème sélectionné revient à contester la description de la situation telle qu'elle existe ou le sens du concept.

L'autre entité extraite d'une phrase est un appel à un schéma. Le cas se produit avec des relatives, les inflexions rattachées aux objets de la phrase. Nous rappelons qu'une inflexion est une phrase à part entière. Aussi le processus lié au traitement d'une phrase recommence dans son intégralité.

Lorsqu'un schéma est rempli par les éléments d'une phrase, nous procédons à deux vérifications.

La première est une vérification de la présence, dans l'occurrence, de la structure "CONDITION". Si elle n'apparaît pas, nous la créons avec tous ses éléments obligatoires.

Nous engendrons ainsi d'autres inférences nécessaires à la compréhension de la situation.

Lorsque nous créons les éléments, nous pouvons savoir à quels objets ils font référence grâce aux procédures qui leur sont associées. Elles précisent les concepts du schéma qu'il faut reprendre. S'ils n'ont pas été instanciés, ou si les éléments contiennent de nouveaux concepts, nousinstancions les valeurs par défauts. En cas de non existence de celles-ci, nous gardons les concepts génériques; ils pourront être précisés par la suite.

Les inférences peuvent porter aussi bien sur des variables que sur des schémas. Elles représentent une information plus concrète que dans le cas exposé plus haut. Ce ne sont plus les éléments d'un raisonnement, mais des points qui doivent se réaliser ou être réalisés, sinon la situation, thème du discours, ne peut avoir lieu.

La seconde vérification se passe à l'intérieur de chacune des structures remplies. Nous avons vu que leurs éléments peuvent être ordonnés. Il s'agit alors de vérifier que si un élément est rempli, ses prédécesseurs

le sont aussi. S'ils ne le sont pas, nous les créons selon le processus décrit ci-dessus.

Nous n'avons décrit ici que les opérations propres à notre utilisation des schémas. Nous n'avons pas cherché à mettre au point une représentation complète des connaissances, avec les procédures concernant d'autres utilisations de ces schémas, comme nous n'avons pas cherché à représenter d'autres connaissances que celles qui nous sont utiles.

II.3.2- Organisation des schémas

1) La structure et sa signification

Nous avons vu qu'un schéma peut être décrit par référence à d'autres schémas. Ces références créent un ensemble de liens et une organisation des schémas les uns par rapport aux autres. Cependant elles ne sont pas le seul type de relation entre deux schémas.

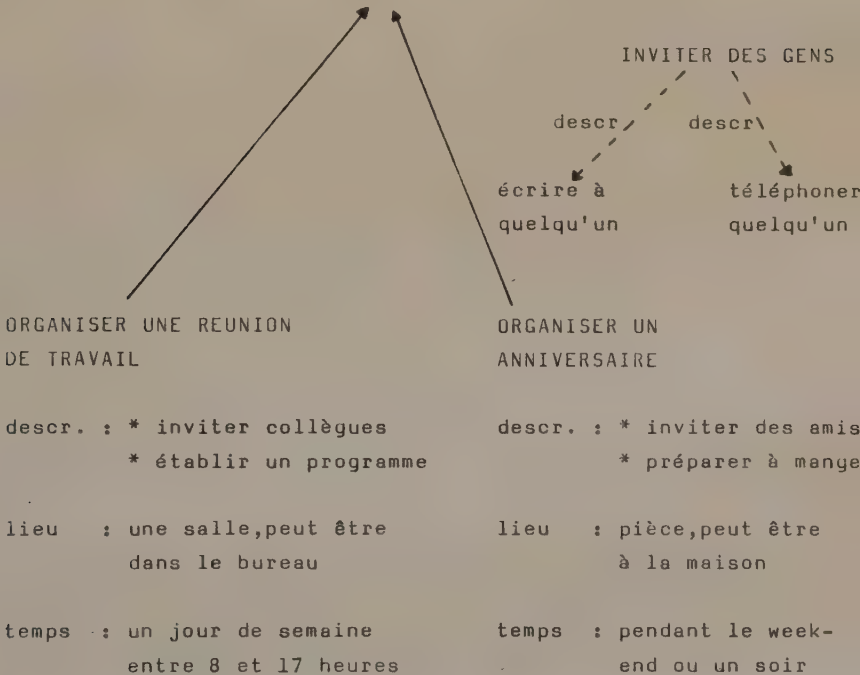
Nous avons créé une structure hiérarchique de schémas au moyen d'une arborescence. En effet, certaines situations sont des cas particuliers d'autres situations.

Par tout cet ensemble de liens, nous pouvons connaître le rapport qu'il y a entre deux schémas, s'il existe.

Cette organisation nous permet de répartir l'information et de posséder une structure modulaire.

ORGANISER UNE REUNION

description: * inviter des gens
temps : doit être une date et
une heure



Les "*" indiquent la référence à un schéma.

Exemple II.6- Relations entre les schémas

Les schémas "organiser une réunion de travail" et "organiser un anniversaire" sont tous deux descendants du schéma "organiser une réunion". Ils en constituent des cas particuliers.

Outre le fait que cette structure est plus économique en mémoire car elle évite des redondances, elle a une signification plus profonde.

Les "frame,script ..." sont sensés donner une image de la mémoire et de son fonctionnement. Ils représentent un "chunk", c'est-à-dire un élément regroupant tout ce qui vient en mémoire quand on parle d'une situation. Mais il est difficile de définir cette information.

Pour Charniak, c'est une liste d'événements classés par ordre chronologique; pour Schank, c'est une liste causale. Mais tous deux cherchent à décrire une situation de A à Z, sans omettre un seul détail.

De plus, toutes les situations ne peuvent être exprimées à l'aide de l'un ou l'autre critère uniquement. Les travaux de Evens (18) ont d'ailleurs montré l'arbitraire sinon l'impossibilité d'une telle démarche.

Au moyen des "MOPS"* , Schank décrit une autre structure (41), où ce que l'on sait sur un sujet ("general event"), se construit à partir de faits particuliers qui arrivent aux gens et dont ils se souviennent ("event memory").

* MOP: Memory Organization Packets

En fait, si beaucoup sont d'accord avec cette idée de modèle ou de stéréotype, aussi bien en psycholinguistique qu'en intelligence artificielle, on ne sait pas exactement les connaissances que cela recouvre. Ceci nous a conduit à regrouper les connaissances d'une autre manière.

Au lieu d'essayer de tout dire au même endroit, nous avons choisi de ne regrouper que l'essentiel d'une situation, c'est-à-dire le minimum de ce qu'on peut en dire pour la décrire de façon compréhensible. Cette notion d'essentiel et de minimum reste évidemment très subjective et a sans aucun doute une valeur différente d'un individu à l'autre.

Nous répartissons ainsi l'information en niveaux de précision différents. Il est en effet inutile de connaître tous les détails au même moment; ils ne font qu'embrouiller la compréhension d'une situation. Cette division des situations en degrés de précision différents nous a semblé respecter la façon dont les individus développent un thème et nous permet de le suivre.

Comme nous l'avons mentionné § 1.4, ils peuvent en effet introduire un sujet de façon générale et le préciser ensuite, ou bien le texte garde un même niveau d'abstraction ...

L'organisation des informations que nous avons choisie nous permet de savoir quel aspect du thème est traité dans le texte étudié. Les termes employés dans

chaque phrase nous indiquent le niveau de détails et permettent d'accéder à un schéma respectant le degré de précision désiré par l'auteur.

Nous sommes donc à même de sélectionner l'information au fur et à mesure des besoins. Celle-ci ne constitue que l'information utile à chaque pas de l'analyse. L'approfondissement de la compréhension demandera certainement plus de connaissances. Il sera possible de les rechercher à ce moment là.

De la même façon, si on pose des questions sur le texte, il est inutile de chercher à y répondre par avance et de rassembler toutes les connaissances que l'on possède sur un sujet. La question précisera le point du sujet à développer.

La représentation mise en place ne concerne pas seulement l'analyse de texte. Elle peut être utilisée par la machine pour intervenir dans un dialogue. Nous faisons référence au travail développé actuellement par A. Vilnat (47).

En effet pour répondre à des questions, ou en poser, le système doit posséder de nombreuses connaissances. La première, évidente, est la connaissance du thème général et de son développement.

Mais surtout, il faut pouvoir faire la différence entre des banalités et des détails, tous deux pouvant être ennuyeux et inutiles s'ils n'intéressent pas l'interlocuteur.

La structure que nous avons mise en place permet de savoir si on est en relation avec un quidam ou un spécialiste, dans la partie compréhension du dialogue; elle permet aussi de répondre en fonction des connaissances de l'interlocuteur, à un niveau équivalent.

Par cette représentation des connaissances sous forme de schémas et par l'organisation de ces schémas entre eux, nous traduisons trois idées essentielles:

- donner un nom précis à ce que l'on décrit, d'où le couple prédicat + objet nommant le schéma. On ne peut sélectionner le schéma au vu du seul objet dans une phrase. Il est indispensable de savoir sous quel point de vue il est abordé pour trouver le contexte auquel la phrase se rapporte.
- organiser les connaissances en niveaux de compréhension.
- sélectionner uniquement l'information pertinente. On évite, par ce graphe de schémas, de prendre en compte des informations inutiles.

Nous cherchons aussi à réaliser une représentation souple, de façon à pouvoir facilement mettre les connaissances à jour et surtout de façon à pouvoir les utiliser simplement.

Cette description ne serait pas complète si nous n'exposions pas les principales caractéristiques de ces

schémas quant à leur utilisation, afin de pouvoir les comparer avec des systèmes existants.

Nous allons les exposer de manière brève. Tous les points concernant directement l'analyse du thème seront développés ou complétés au chapitre suivant.

2) Quelques points sur l'utilisation de la base de connaissances

Le premier point quant à l'utilisation de ce graphe est le plus important: nous n'exploitons pas cette structure afin de produire des inférences dès qu'un texte mentionne une des situations décrites.

Nous avons vu que l'analyse que nous faisons d'un texte est à la fois montante et descendante. Par l'organisation des connaissances mise en place, nous respectons tout à fait cette idée. Toutes les relations existant entre les schémas permettent une analyse montante, dans le sens où nous allons chercher un chemin à partir d'un schéma sélectionné par une phrase vers un groupe d'autres schémas: le contexte.

Toujours à propos de l'analyse du texte, notre but est de mettre en évidence les liens entre chaque phrase et de définir leur nature. Grace aux liens hiérarchiques et aux références à d'autres schémas pour décrire un schéma particulier, références liées à une structure possédant un nom (description, condition, résultat, temps, lieu), nous sommes capables de résoudre ces problèmes.

Lorsque nous avons trouvé le lien entre deux schémas, nous en connaissons alors la nature. Il est indiqué soit par le nom de la structure, soit par le fait que l'un est père, ou fils, de l'autre. Ce nom est reproduit dans l'occurrence du schéma traité.

Au niveau de l'implantation physique, il est simple de passer d'un schéma à un autre et de connaître la raison de ce lien.

Il nous faut préciser un point à propos de la structure hiérarchique. Il n'est jamais question d'héritage entre schémas au sens où on l'entend habituellement. Dès que l'on trouve un schéma, on ne reporte pas à son niveau toutes les connaissances plus générale que l'on possède. Ceci serait contraire à nos principes de niveau de compréhension et de sélection de l'information pertinente.

S'il est nécessaire de connaître des généralités sur le thème, nous irons les chercher à ce moment car elles feront alors partie du raisonnement à trouver.

La notion de hiérarchie est introduite surtout parce qu'elle correspond à une réalité aussi bien pour ce qui est du rapport de certaines informations entre elles que pour la façon de traiter un sujet.

Nous avons présenté le problème de certaines métaphores au chapitre précédent. Il s'agit de trouver le sens d'une phrase lorsque l'analyste ne parvient pas à choisir les concepts correspondant aux mots de la phrase à cause du non respect de contraintes sémantiques.

Ce problème pourra être résolu grâce à la représentation hiérarchique des situations. En effet quel est le rapport entre deux schémas situés chacun sur une branche d'un arbre* ?

De tels schémas décrivent des situations qui ont la même signification à un niveau plus élevé. Elles ont même sens général et diffèrent par certaines particularités; aussi les prédicats qui les décrivent peuvent-ils être considérés comme synonyme dans le contexte précisé par le schéma commun, par exemple soigner et réparer dans le contexte de remettre en état.

II.3.3- Comparaison avec des structures de données existantes

Parmi les principales caractéristiques de notre travail, nous avons insisté sur la réalisation d'une analyse montante et la sélection des informations au fur et à mesure des besoins.

Ces deux principes sont en complète opposition avec ceux développés par Schank au moyen des "scripts" - que nous traduisons par schémas - et des "plans".

Un schéma, tel que le décrit Schank, représente une histoire complète; lors de l'analyse d'un texte, on essaye d'abord de sélectionner le bon schéma, et ensuite, de le remplir complètement.

On retrouve le même principe dans la dernière

* Nous rappelons que dans le cas des seules relations hiérarchiques, nous sommes en présence d'une arborescence.

représentation élaborée par Schank, les "MOPS" et les "TAUS"*; même si les connaissances sont mieux réparties, de façon plus modulaire.

Ce type de représentation retraçant une histoire complète, avec tous les enchaînements prévus entre les différentes actions, implique de se limiter à un domaine restreint.

C'est ainsi que Boris, Dyer (16), Lehnert (32), programme qui utilise les "MOPS" pour analyser des textes dont le thème est un divorce, est amené, au cours de son analyse, à se poser la question suivante: Que peut-on faire dans une chambre? La réponse donnée par l'intermédiaire des "MOPS" et des scénarios est: dormir, changer de vêtements et avoir des relations sexuelles (32 p.40).

Ce type de description est trop influencé par l'histoire que l'on veut analyser. Il est certain que cette méthode permet de trouver les raisons complètes de toutes les actions des personnages. On peut ainsi faire des inférences très précises à partir de chaque phrase. Par contre, on pourra difficilement utiliser ces connaissances pour des histoires non prévues.

A ce niveau, ce ne sont pas tant les structures qui sont en cause, que leur contenu et leur utilisation. Ces deux éléments constitueront les points de comparaison entre les travaux de Schank et le nôtre.

Afin de ne pas se limiter à un type d'histoires sur un sujet, nous avons choisi de dissocier les

* TAU: Thematic Affect Unit

connaissances de situations courantes et les thèmes permettant de les assembler pour en faire une histoire.

Un des problèmes de Schank est la sélection d'un schéma et la confirmation de ce choix. De façon extrême, cette méthode rend presque inutile la lecture de l'histoire après confirmation: on est capable de la deviner. On essaye d'ailleurs de le faire après chaque phrase en produisant de nombreuses inférences.

Nous avons choisi, nous, de sélectionner les schémas au cours de l'histoire, la connaissance du thème et de son déroulement permettant de reconstituer le raisonnement. Nous construisons ainsi une représentation de l'histoire au cours de l'analyse, et celle-ci dépend de l'auteur, de sa façon d'enchaîner ses idées.

Nous n'essayons jamais de deviner ce qui peut suivre et nous ne partons pas d'un comportement et d'une situation déclarés normaux afin de pouvoir prévoir et répertorier tous les cas anormaux.

Schank regroupe dans ses schémas trois éléments:

- la représentation du thème et de son développement;
- la représentation d'une histoire;
- la signification du non respect d'une prévision.

A l'opposé, nous avons choisi de séparer ces connaissances. Ainsi les problèmes se posent différemment.

Le problème de la sélection ne se pose plus pour nous.

En effet, pour trouver le bon schéma, Schank sélectionne les schémas contenant un terme choisi de la phrase et attend la confirmation de l'un d'eux par les concepts suivants.

En dissociant les connaissances, nous avons supprimé le problème. Un schéma, une fois sélectionné par son nom, n'est pas remis en cause. Quant à savoir s'il constitue une confirmation du schéma précédent, ou s'il est relié à d'autres schémas, cela relève du suivi du thème.

Nous n'avons pas non plus à multiplier les schémas décrivant des histoires pourtant analogues. Il ne nous faut pas prévoir toutes les histoires dans leur moindres détails et complications, mais de nombreuses situations simples, décrites de façon minimale, qui, regroupées, traduiront une histoire.

Schank a défini plusieurs types de schémas:

- les schémas de personnage ("personal script");
- les schémas instruments ("instrumental script");
- les schémas de situations ("situational script").

Les premiers permettent de décrire une série d'actions qu'un personnage, possédant un trait particulier, accomplit pour atteindre un but (ex: le comportement d'un bon samaritain).

Les deuxièmes, très nombreux, décrivent des séquences d'actions très rigides qui se réaliseront toujours de la même façon. Ils ne constituent pas des

sujets d'histoire (ex: allumer une cigarette, démarrer une voiture).

Les derniers sont des sujets d'histoire où on essaye de prévoir tout comportement anormal.

Si on compare nos schémas à ceux de Schank, il y a plus d'analogies avec les "scripts" du deuxième type qu'avec les autres. En effet ces derniers servent à décrire les histoires; par contre ils ne jouent pas un très grand rôle. Nos schémas aussi servent à cela. Mais ils sont conçus de manière beaucoup plus souple.

L'ordre des événements à l'intérieur d'un schéma ne tient pas uniquement à une relation causale entre eux. Il peut s'agir de relations temporelles; il peut aussi n'y avoir aucun ordre entre deux événements.

En fait notre méthode s'apparente à la construction du schéma de situation ("situational script") pour chaque histoire.

Nous avons vu que depuis les schémas, Schank a élaboré un autre modèle de représentation. Celui-ci provient d'une théorie sur l'acquisition des schémas et sur leur représentation en mémoire. Les "MOPS" sont des super schémas. Ils permettent de mieux répartir les connaissances et de ne pas répéter la même chose dans des schémas différents. Ces derniers représentent maintenant la généralisation d'événements.

Si la représentation des connaissances à l'aide de "MOPS" est différente, mieux organisée qu'avec les an-

ciens schémas, l'utilisation qui en est faite repose sur les mêmes principes. Ils contiennent aussi les mêmes connaissances.

Pour conclure, nous pouvons ainsi résumer les principales différences entre les systèmes de Schank et Lehnert et le nôtre:

- méthodes d'analyse opposées.

Tandis que Schank et Lehnert développent des analyses descendantes et recherchent toutes les inférences possibles, nous appliquons une analyse montante et descendante qui permet de ne produire que les inférences indispensables.

- nous séparons la représentation des connaissances suivantes:

- * représentation du thème
- * représentation de son déroulement
- * représentation de l'histoire, qui est en fait construite au cours de l'analyse

Ces connaissances sont au contraire regroupées dans les systèmes cités.

Nous allons nous intéresser maintenant aux travaux de Charniak.

Que ce soit dans sa thèse (12) ou dans son programme Mrs Malaprop (14), Charniak a conservé les mêmes principes en ce qui concerne la production d'inférences.

Il n'est pas question, selon lui, de faire toutes les inférences possibles à partir d'un fait. Aussi il a

cherché à les limiter par le contexte.

Dans sa thèse, son but est de comprendre des histoires pour enfants et de répondre à des questions sur ces histoires.

Au lieu de faire toutes les inférences possibles pendant la lecture des phrases, il essaye de ne les engendrer que lorsque cela est nécessaire pour comprendre l'histoire (comprendre les référents, le pourquoi des actions) ou pour répondre aux questions.

Pour cela, il crée les "démons", attachés à certains concepts: les concepts thèmes ("topic concept"). Un démon est une règle qui se déclenche lorsque le concept auquel elle est associée intervient dans l'histoire de façon explicite ou implicite. Les démons représentent ainsi des connaissances très ponctuelles, reliés à des concepts, et donc indépendants de l'histoire, du contexte, dans laquelle ils apparaissent.

Ex: Si il pleut (ou va pleuvoir) et si une personne P est dehors, alors P sera mouillé.

Cette règle est associée au concept thème: pleuvoir.

Si un démon ne peut s'appliquer au moment de son déclenchement, il reste en éveil jusqu'à ce qu'il obtienne l'information désirée. Les démons sont donc capables de faire des inférences en avant et en arrière.

Les démons pouvant rester en éveil, cela pose le problème de leur contrôle. Quel critère appliquer afin

de savoir si la règle est toujours en rapport avec l'histoire ou si elle est devenue caduque ?

Un autre inconvénient provient du nombre des démons qu'il faut prévoir pour tous les concepts.

Avec un tel procédé, Charniak va quand-même produire des inférences inutiles, et risque même de déduire des informations erronées. Ce type de processus découle de ce que les démons ne sont pas reliés entre eux et dépendent de l'apparition d'un concept.

Ex: Le démon D1 est rattaché à pleuvoir. Il est déclenché par la possibilité de pleuvoir.

D1: Si une personne peut être prise sous la pluie, elle prend un parapluie pour ne pas être mouillée.

Cette règle n'est pas activée à l'énoncé de la phrase suivante:

Jack commença à s'inquiéter quand il réalisa que tout le monde dans la rue avait un parapluie.

Pour comprendre cette phrase il est nécessaire de relier des connaissances entre elles et ceci ne peut être réalisé au moyen de règles. Cela requiert l'élaboration de structures de données complexes.

Aussi, nous n'avons pas voulu réaliser ce type de processus. Aux démons de Charniak, nous avons substitué des schémas* qui sont déclenchés uniquement quand ils sont nécessaires à la compréhension du discours, donc en fonction du thème développé.

* Pour une étude comparative des démons et des schémas, on peut se reporter à l'article de Charniak (13, p. 152-153).

Charniak s'est ensuite tourné vers les schémas puisque, dans son programme Mrs Malaprop, il décrit un scénario de peinture, comprend des phrases et répond à des questions sur ce sujet. Son schéma est construit de façon modulaire, dans le sens où il utilise des informations déduites d'autres schémas. Nous suivons d'ailleurs tout à fait Charniak sur ce point.

Dans son scénario, les événements sont ordonnés chronologiquement. Ils indiquent la marche à suivre pour repeindre un objet, et cela de façon extrêmement détaillée.

Ex: L'objet à peindre doit être propre
 provient de laver un objet
 mettre du papier sous l'objet
 répéter
 mettre de la peinture sur l'instrument
 mettre l'instrument au contact de l'objet
 ...

A pratiquement chaque étape, on fait appel aux résultats d'autres scénarios, tout aussi détaillés.

Nous avons déjà signalé l'inconvénient d'un tel ordonnancement. Nous allons développer les problèmes posés par le contenu d'un scénario.

Le nombre de connaissances qui y sont regroupées nous semble en effet être le point le plus critiquable de cette représentation. Nous ne reviendrons pas sur l'inutilité de regrouper tous les détails d'une

situation en une même structure ni sur les inconvénients que cela entraîne. Nous avons déjà traité ces sujets.

Il nous faut néanmoins préciser que décrire ainsi une situation provoquera une description tout aussi abondante d'une situation voisine, par exemple une autre manière de peindre, ou de changer la couleur d'un objet, alors qu'elles possèdent toutes des points communs. De plus, ces similitudes ne sont pas exprimées au travers de ce type de scénario alors qu'il est parfois important de les connaître.

Les modèles de représentation élaborés par Schank et Charniak sont destinés à une compréhension approfondie de certaines histoires. Ils ont représenté plus de connaissances que celles que nous possédons dans nos schémas. C'est pourquoi nous avons limité les comparaisons aux connaissances équivalentes dans les différents systèmes.

D'autres ont réalisé des programmes utilisant des schémas: GUS, Bobrow (7) et NUDGE, Goldstein (22). Ceux-ci constituent en fait des applications d'un langage de représentation, KRL pour GUS et FRL, Bobrow (6) pour NUDGE.

GUS est tout à fait similaire à SAM aussi nous ne jugeons pas nécessaire une nouvelle comparaison avec notre système.

Nous allons un peu plus étudier FRL, mais pour la structure de ses schémas, et non pour les connaissances qu'ils contiennent.

Un schéma contient une liste de variables. A chacune d'elles sont associées plusieurs facettes. Les unes précisent des valeurs, d'autres des valeurs par défaut. Il est aussi possible de préciser par ce moyen des contraintes et des préférences sur les valeurs devant remplir les variables.

Des procédures sont également attachées à chaque variable. Celles-ci sont exécutées lorsqu'une valeur est donnée à la facette (procédure "IF-ADDED"), ou si on a besoin d'une valeur ("IF-NEEDED").

Ils ont aussi créé une structure hiérarchique de schémas, qui a la même signification que la hiérarchie dans un réseau sémantique.

Nous n'avons pas reproduit exactement ces caractéristiques, mais nous nous en sommes inspirés pour élaborer la structure d'un schéma.

Le modèle de représentation des connaissances que nous avons établi essaye de respecter trois principes importants:

- donner un nom représentatif à ce que l'on décrit
- organiser les connaissances en niveaux de compréhension
- sélectionner uniquement l'information pertinente

Comme nous l'avons déjà mentionné, nous n'avons pas cherché à représenter toutes les connaissances nécessaires à la compréhension d'une histoire.

Nous avons voulu établir une structure générale de ce modèle et indiquer la répartition des connaissances à l'intérieur de ses composants.

Chapitre III

Description des principales procédures

III.1- Recherche du thème principal

Nous avons vu §I.2 que la recherche du thème principal repose sur un principe général : lorsqu'une personne veut parler d'un sujet, elle l'introduit par quelques phrases. Nous allons exposer en détail le processus de sélection de ce thème.

- On émet l'hypothèse que le premier schéma déclenché indique le thème principal.

On sélectionne avec lui l'ensemble des schémas qui le décrivent dans notre graphe: le contexte.

Cette sélection de schémas non encore déclenchés pour constituer le contexte est propre à la recherche du thème principal. Nous considérons qu'un seul schéma n'est pas suffisant pour représenter ce thème. Il correspond à une notion plus vaste, et doit comprendre l'ensemble des sujets qui le composent.

Le nom d'un thème, nous l'avons vu, est le nom d'un schéma. A chaque étape du processus, le nom du thème est celui du schéma qui a permis de constituer le contexte et pour simplifier, nous désignons souvent le thème par

ce schéma.

Nous allons chercher à confirmer ou à modifier notre hypothèse de thème principal.

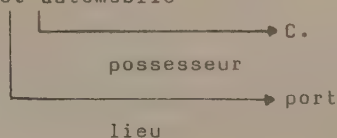
- Nous allons essayer de trouver le rapport entre la phrase suivante et le thème supposé.

Selon sa nature, cette phrase correspond soit à un schéma, soit à une variable. Nous allons examiner le processus mis en place pour chaque cas.

* cas d'une variable

Ex: C. a un canot automobile dans le port.

Représentation interne: canot automobile



La phrase précise un concept, "canot automobile", par deux de ses qualités. Nous considérons que l'emploi du verbe avoir est neutre et ne constitue pas le qualificatif de la phrase. Ces verbes sont reconnus par le fait qu'ils ne constituent pas des entrées vers des schémas.

Si on introduit ainsi une variable sans indiquer le point de vue sous lequel on l'aborde, nous supposons que ce dernier est précisé dans le contexte. Aussi nous allons la rechercher dans les schémas lui appartenant.

Nous la recherchons aussi bien dans les noms de ces schémas que dans leurs structures.

Si la recherche aboutit, c'est la confirmation du thème. Cette confirmation peut avoir lieu de deux façons:

- la variable appartient à une structure du thème principal, ou au nom de l'un des schémas du contexte;
- la variable appartient à l'une des structures des schémas décrivant le thème principal.

Dans le premier cas, le thème actif est aussi le thème principal.

Dans le second cas, il y a déviation et le thème actif devient le schéma contenant la variable. C'est une manière de détailler l'un des points du sujet.

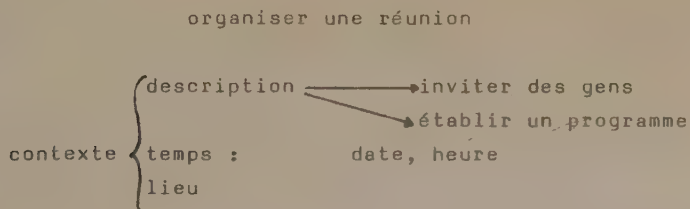
Au contraire si la variable ne fait partie d'aucun schéma, nous la conservons. Nous chercherons à la rattacher au prochain schéma appelé. Nous supposons que cette nouvelle variable précise un élément d'un thème qui sera développé par la suite. Le terme peut aussi correspondre à un sens inconnu du système, ou à une métaphore (nous en verrons un exemple au chapitre suivant).

Ce cas peut se produire avant la sélection d'un premier schéma. Nous agissons alors de la même façon.

Je désire organiser une réunion; ce sera à 16 heures.
Nous nous réunirons dans mon bureau.

Exemple III.1- Confirmation du thème principal

La première phrase provoque le déclenchement du schéma "organiser une réunion". Il représente alors le thème supposé du discours, avec les éléments qui lui sont liés.

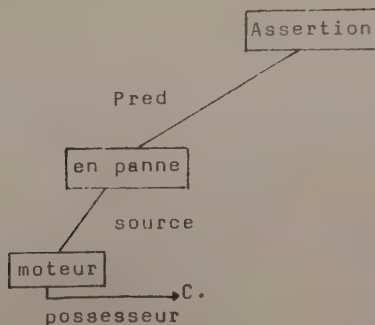


La seconde phrase permet de donner une valeur à l'heure; il y a donc confirmation du thème.

* cas d'un schéma

Le moteur de C. est en panne

Représentation interne



Le schéma appelé est : moteur en panne.

Si ce schéma fait partie du contexte, c'est la confirmation de notre hypothèse. Dans le cas contraire, nous recherchons dans le graphe "modèle", s'il existe un chemin qui relie ce dernier schéma à l'un des éléments contextuels. Si ce chemin existe, c'est encore la confirmation du thème, dans le sens où nous interrompons la recherche du thème principal. Mais nous allons voir que le thème principal peut être différent de l'hypothèse formulée.

Lorsque nous cherchons le thème du discours, nous préférons choisir un thème qui soit le plus général et le plus représentatif de l'histoire. C'est pourquoi, il dépend du chemin d'inférences trouvé. Un tel chemin est trouvé grâce aux relations inter-schémas.

Les relations recensées à partir d'un schéma sont de quatre types:

- 1) le lien reliant le schéma à son père;
- 2) les liens le rattachant à ses fils;
- 3) les liens des schémas le décrivant, trouvés dans ses éléments;
- 4) les liens des schémas qu'il décrit.

Les liens de type 1) et 2) entre deux schémas indiquent la situation la plus générale. Les liens de type 3) et 4) traduisent des niveaux de dépendance entre les schémas.

Ces liens sont regroupés en deux classes. La première est constituée des liens 1) et 4). La seconde est formée des deux autres liens. Nous expliciterons la signification de ces classes plus loin.

Nous partons du schéma représentant le thème principal et nous parcourons le chemin pour rechercher s'il existe une relation de la première classe parmi les inférences. Nous trouvons ainsi le thème le plus représentatif de l'histoire. Nous allons exposer pourquoi nous préférons choisir un thème amené par un lien 1) ou 4).

Il est inutile de revenir sur la signification d'un lien n°1.

Nous considérons qu'un schéma introduit par une relation n°4 représente une situation mieux adaptée à l'histoire. En effet un schéma trouvé par un lien n°3 dépend fortement du schéma qui l'appelle puisque certaines valeurs de ses éléments proviennent de ce schéma.

Si nous trouvons un tel schéma, il devient le thème principal du discours. Sinon nous confirmons le thème supposé.

Le moteur est en panne.

Marius est garagiste.

Exemple III.2- Confirmation du thème.

che dans le graphe n'a pas abouti, nous sommes alors en présence d'un autre thème et le thème principal devient le dernier schéma déclenché.

Il nous faut quand même rechercher si l'enchaînement est cohérent. Nous allons essayer de trouver un lien entre les deux schémas : celui appelé par la phrase, et celui représentant le thème supposé.

Ce lien existe si les deux schémas utilisent un même concept. Cela signifie que la première phrase - à l'origine du thème principal - a pour rôle de préciser ou de situer l'un des éléments du thème évoqué dans la phrase suivante. Nous détaillerons cette recherche dans la partie vérification de la cohérence.

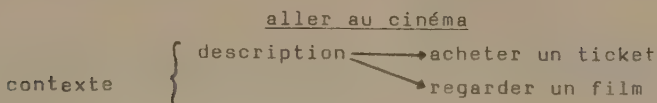
Si le thème n'a pu être confirmé, nous recommençons le processus avec le nouveau schéma comme hypothèse. Le schéma déclenché auparavant est annulé. Nous conservons seulement le concept lui appartenant, s'il a été trouvé, qui permet de remplir le nouveau schéma. En effet, la phrase précédente ne servait qu'à préciser un élément de ce thème et n'introduisait pas un sujet de discussion.

Si l'on veut ensuite reparler du premier sujet, on le fera explicitement, sinon on rendrait cela incohérent pour le receveur du discours.

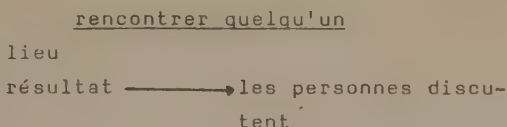
Je suis allé au cinéma hier et j'ai rencontré Jean.
Il m'a dit qu'il s'était acheté une voiture.

Exemple III.3- Changement de thème

La première proposition permet de déclencher le schéma "aller au cinéma".



Avec la seconde proposition, nous appelons le schéma "rencontrer quelqu'un".



Nous cherchons un lien entre ce schéma et le contexte. Nous n'en trouvons aucun. Il s'agit donc d'un changement de thème et le thème principal devient la rencontre avec Jean. La phrase suivante confirme d'ailleurs ce thème.

Quant au lien de cohérence entre les deux phrases, il est dû au concept "cinéma". En effet le schéma "rencontrer quelqu'un" demande obligatoirement un lieu. Nous savons que le cinéma est un lieu (par le réseau sémantique), aussi nous remplissons le lieu de rencontre par "cinéma", tandis que "aller au cinéma" est oublié.

Nous recherchons le thème principal afin d'avoir une référence permettant de situer les autres sujets et de donner ainsi un point de départ à l'analyse.

III.2- Rapport phrase-schéma et vérification de la cohérence.

Une fois le thème principal sélectionné, nous pouvons poursuivre l'analyse et vérifier la cohérence du texte.

Nous avons vu que nous sélectionnons un nom de schéma à partir des couples Prédicat + Objet contenus dans les phrases. Il nous faut préciser que toutes les phrases n'ont pas pour fonction d'appeler un schéma, c'est-à-dire d'introduire un sujet.

Nous considérons que certaines n'ont aucune signification pour notre analyse. Par exemple, une phrase telle que : X m'a dit de m'adresser à M, ne provoque pas l'appel à un schéma. Elle sert en fait à introduire un nouveau personnage, M. Ce qui est intéressant dans une telle phrase, c'est le rapport entre ce personnage et l'histoire. Il faudrait, pour prendre cette information en compte, introduire un traitement particulier des personnages, Berthelin (4). De tels systèmes ayant déjà été réalisés, nous ne nous sommes pas intéressé à ce problème.

Les phrases qui interviennent dans le raisonnement du programme exposent, elles, des situations ou précisent des variables. Une fois reconnues, il nous faut les relier à ce qui est connu.

Possédant le point de départ de la recherche du lien, il nous faut préciser son point d'arrivée. C'est

alors que nous décrivons le contexte, son contenu ne valant qu'à ce moment particulier de l'analyse. Nous avons vu § 1.5 la construction générale du contexte. Nous la détaillerons par la suite, lorsque nous aurons exposé la représentation du déroulement du thème. Précisons seulement que les thèmes y sont ordonnés, le premier étant celui qu'il est plus cohérent de développer.

Nous allons distinguer deux cas. On extrait de la phrase, soit le nom d'un schéma, soit une variable.

* cas d'un schéma.

Rechercher un lien entre le schéma de la phrase et le contexte signifie ici rechercher un lien dans le graphe modèle reliant le schéma à l'un de ceux du contexte.

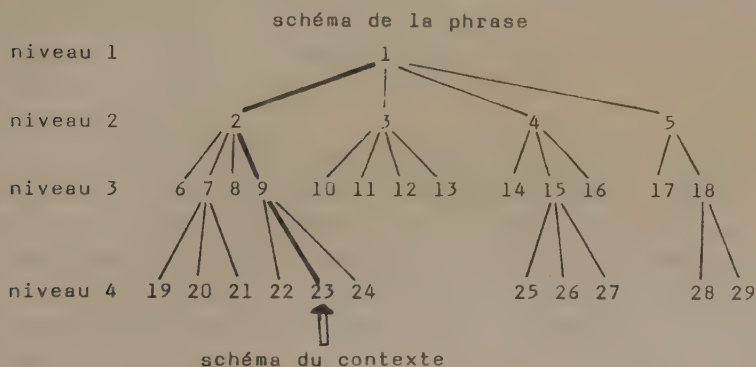
Nous avons recensé les différentes relations permettant de passer d'un schéma à un autre au paragraphe précédent. Nous allons voir comment nous les recherchons.

Nous construisons un arbre pendant cette recherche. Nous le construisons et le parcourons en largeur. Nous créons en une fois tous les descendants d'un noeud: ce sont tous les schémas reliés au même schéma. Nous recherchons alors si l'un ou plusieurs des descendants fait partie du contexte.

Que l'on trouve ou non, nous passons au schéma suivant de même niveau que celui qui vient d'être

décrit, et nous réitérons la même opération.

Lorsque tous les schémas d'un niveau ont été traités et que nous n'avons pas trouvé de chemin, nous descendons dans l'arbre et nous recommençons le même processus à partir du schéma le plus à gauche.



Exemple III.4- Arborescence engendrée par la recherche d'un lien

A partir de 1, nous recherchons la liste des schémas qui lui sont reliés (2,3,4,5); aucun ne fait partie du contexte.

Nous descendons d'un niveau et cherchons la liste des schémas pour 2 (6 ... 9); nous poursuivons la même opération pour 3, puis 4 et 5. Nous passons ensuite au niveau suivant, etc ...

Le chemin trouvé 1,2,9,23 représente le raisonnement nécessaire à la compréhension de la phrase.

Nous construisons cet arbre ainsi, car nous désirons trouver le raisonnement le plus court, donc le plus simple.

Le fait d'avoir ordonné les thèmes dans le contexte joue un rôle à ce niveau. Les sujets y sont donc classés selon un critère de plus forte cohérence. Si l'on trouve plusieurs chemins de même longueur, nous choisissons celui qui aboutit au thème le plus fort.

L'arborescence que nous développons ne peut atteindre de grandes proportions. Nous arrivons assez rapidement aux schémas "primitifs". Lorsque nous arrivons à ce niveau sans avoir trouvé de chemin, il nous faut rechercher un autre type de raisonnement, un autre critère de cohérence. Nous sommes en présence d'un changement de thème.

Un nouveau sujet est introduit de façon cohérente s'il développe, avec une approche totalement différente de celle qui est en cours, un concept introduit explicitement ou implicitement par la phrase précédente. Il s'agit donc de trouver le concept qui appartient au thème actif et qui est repris et développé dans la dernière phrase.

Pour cela, nous nous servons de critères syntaxiques et contextuels. Les premiers viennent d'un éventuel emploi de pronoms.

Si leurs antécédents sont trouvés, nous recherchons un pronom possédant un antécédant situé dans la phrase

précédente et correspondant à un nouveau concept.

Nous sommes dans le cas d'un concept commun explicite. La liste des antécédents est établie lors de la résolution des pronoms.

Il (le moteur) était sur le bateau de M. frère. Vous (César) le (M. frère) connaissez ?

Exemple III.5- Lien dû à un pronom

Le lien peut être moins évident et c'est alors dans le contexte qu'il nous faut le rechercher. Nous traitons ici le cas d'un concept commun implicite.

Nous envisageons plusieurs possibilités.

- Le nouveau schéma contient une structure pouvant être remplie par un des concepts explicite de l'occurrence du thème principal, qu'il soit dans son nom ou dans l'une de ses structures.

Je suis allé au cinéma et j'ai rencontré Jean.

Le lieu de rencontre est le cinéma mentionné dans le thème actif. Il remplit un élément du schéma "rencontrer personne".

On re peint le cinéma Y. J'ai vu un film super hier.

"cinéma" permet aussi de remplir un élément du schéma "voir un film". Ce dernier type d'enchaîne-

ment correspond plus à une association d'idées, mais il est très difficile de faire la différence entre ces deux derniers raisonnements à ce niveau de l'analyse.

- Un concept explicite du nouveau schéma fait partie du thème actif. Ce concept est l'un de ceux que précise la dernière phrase. Le lien est alors moins immédiat.

(On sait que Jeanne est la femme de Marius)

J'ai rencontré Marius. Jeanne a acheté une voiture.

Le nouveau schéma "acheter une voiture" ne peut être rempli par un élément de "rencontrer Marius". Nous recherchons alors un lien entre les concepts de la phrase "Jeanne" et "voiture" et les concepts du thème actif. La cohérence du texte provient de ce qu'il existe une relation entre Jeanne et Marius.

Nous venons d'exposer le traitement de vérification de la cohérence lorsque la phrase permet d'appeler un schéma. Nous allons maintenant nous intéresser aux phrases précisant une variable.

* cas d'une variable

Il ne s'agit pas ici de rechercher un chemin dans le graphe. Si une variable est précisée, sans que l'on donne le point de vue sous lequel on l'envisage, nous supposons que ceci a déjà été fait et que cette variable peut se rattacher directement au contexte.

Nous recherchons donc si les thèmes faisant partie du contexte contiennent cette variable dans l'un de leurs éléments. Si nous ne la trouvons pas, nous optons pour un changement de thème. La variable permet d'introduire l'un des éléments du nouveau sujet, qui sera précisé dans la suite ou alors nous ne connaissons pas le sens exact de ce concept (sens inconnu, métaphore).

III.3- Représentation du texte et du déroulement du thème.

III.3.1- Représentation du texte.

Nous avons vu les différents enchaînements pouvant exister pour intégrer une phrase à ce qui a déjà été analysé. Il nous faut alors conserver ces informations, tant celles qui détaillent un sujet, que celles qui expliquent un raisonnement.

Nous avons déjà mentionné que nous séparons la représentation du texte de celle du déroulement du thème. Nous allons donc commencer par la représentation du texte, constituée par le graphe des occurrences de schémas.

Nous rappelons que ces schémas expriment soit un sujet effectivement traité, soit un point d'un raisonnement.

Nous reproduisons entre les schémas les seuls liens qui ont été mis en évidence par les "chemins de cohérence".

Dans le cas des changements de thème où le lien a été trouvé, nous construisons une nouvelle relation à partir du nouveau sujet vers celui auquel il se rattache. Nous y précisons le concept commun.

Nous n'avons pas développé l'instanciation d'un schéma car elle dépend fortement des référents trouvés. En effet si l'on connaît déjà une occurrence du schéma étudié, ce sont les référents qui indiquent s'il s'agit du même schéma ou si l'on reprend le même thème, mais en parlant de nouveaux objets. Ceci ne modifie en rien la cohérence du texte puisque le thème développé est cohérent avec ce que l'on connaît du discours.

III.3.2- Représentation du suivi du thème et mise à jour du thème actif.

Nous représentons le déroulement du thème au moyen d'une arborescence. Le noeud origine de cet arbre est le thème principal.

Entre deux noeuds, deux thèmes, nous n'indiquons pas la nature du raisonnement qui a permis de passer de l'un à l'autre. Ceci est déjà conservé dans la représentation du texte. Nous indiquons ici la nature du déroulement, à savoir s'il s'agit d'un changement de niveau de compréhension, d'une déviation, d'une inférence ou d'un changement de thème.

Après l'analyse de chaque phrase, nous cherchons le thème actif à ce moment du texte. Pour cela nous analysons le chemin permettant d'intégrer la phrase au contexte.

Lorsqu'il s'agit de traiter une variable, trois cas peuvent se présenter.

- confirmation d'un thème du contexte.

On remplit alors le schéma par la variable et on repositionne le thème actif au thème confirmé.

- déviation du thème: la variable remplit une structure d'un schéma décrivant un thème du contexte.

Ce schéma devient le thème actif et nous le rattachons au thème qu'il décrit par un lien indiquant la déviation.

- changement.

Nous attendons la phrase suivante et nous conservons la phrase en mémoire. Rien n'est modifié.

Les cas sont similaires lorsqu'il s'agit de relier un schéma au contexte.

- confirmation d'un thème du contexte.

On remplit le schéma confirmé par le dernier schéma déclenché.

S'il s'agit d'un approfondissement ou d'une généralisation, nous considérons que ce changement de niveau doit être pris en compte. Aussi, dans ce cas, le thème actif est le thème évoqué par la phrase et nous le

reliions au thème du contexte par un lien indiquant un changement de niveau.

Si la confirmation est prouvée par d'autres liens, le thème actif est le thème appartenant au contexte.

- déviation.

Nous séparons deux cas: la longueur du chemin est 2, nous n'avons fait qu'une inférence, et la production de plusieurs inférences.

1er cas: nous cherchons si le schéma représentant l'inférence a déjà été instancié dans le thème auquel il est relié. Si oui, on parle d'un sujet déjà introduit. Il ne s'agit plus d'une inférence. Nous considérons que la phrase développe ce thème. Nous indiquons qu'il y a déviation d'un sujet.

Si le schéma déclenché par la phrase constitue une généralisation ou un approfondissement, on garde le même sujet mais à un niveau de compréhension différent. Le thème actif est alors ce schéma.

S'il n'y a pas de changement de niveau, le thème actif est alors le sujet dévié. Le schéma déclenché par la phrase lui est relié. Il est considéré comme une confirmation de ce thème.

2ème cas: nous créons des occurrences afin de représenter les inférences. Nous indiquons que c'est

par une inférence que nous avons pu passer d'un thème à un autre. Le thème actif est le thème évoqué par la phrase.

- changement.

Le thème actif est le thème sélectionné grâce à la phrase. Nous le relient à l'ancien thème actif en indiquant qu'il y a changement.

Dans le cas d'une incohérence - nous n'avons pas trouvé de lien entre le thème actif et le nouveau schéma - nous ne relient pas l'occurrence de ce dernier schéma à celles qui représentent le texte. On ne sait pas passer de l'un des sujets déjà traités au nouveau thème.

III.4- Construction du contexte.

Le contexte regroupe les sujets en cours de discussion: ce sont les thèmes qu'il est cohérent de développer dans la suite du texte. Nous avons vu que chaque phrase permet de mettre à jour le thème actif; il dépend du raisonnement trouvé.

Cela signifie que les sujets appartenant au contexte sont ré-évalués après chaque phrase. En effet le thème actif influence la composition du contexte.

Nous avons envisagé des démarches qui conservent sa cohérence au discours, à condition que ces démarches, qui partent de la phrase analysée, aboutissent à un

ensemble défini de sujets: le contexte.

Le choix de ces sujets dépend du thème actif, du déroulement du thème pour arriver au thème actif et enfin, du thème principal.

Nous rappelons que les thèmes formant le contexte sont ordonnés par ordre de plus forte cohérence.

Ainsi, le contexte peut être constitué des sujets suivants, par ordre de préférence:

- le thème actif au moment où arrive la phrase;
- le thème qui était actif avant celui-ci, s'il n'y a pas eu changement de thème entre les deux;
- s'il y a eu changement de thème entre le thème actif et le thème principal, le thème qui était actif au moment du changement;
- le thème principal.

Que signifient ces choix ?

Le thème actif est évidemment préféré à tous les autres. Si la phrase se rattachait à ce sujet, on continuerait à parler du même sujet, que ce soit en le confirmant ou en le déviant.

Le deuxième sujet cité signifie que l'on peut revenir à un thème qui vient d'être abordé après en avoir précisé un point. Nous considérons que ce thème est suffisamment proche du thème actif pour pouvoir être repris même sans qu'il en soit fait explicitement mention.

Pour le troisième thème possible, cela correspond à fermer une parenthèse pour revenir au sujet discuté avant celle-ci.

Nous avons déjà dit que l'on peut toujours revenir au thème principal, à n'importe quelle place du discours. Aussi celui-ci constitue-t-il l'élément permanent dans le contexte.

Des exemples sont donnés dans le chapitre suivant.

Nous avons détaillé la procédure de recherche d'un "chemin d'inférences". Définir ainsi un contexte ne signifie pas que l'un des sujets va être directement développé.

Cela signifie qu'il est indispensable de relier la phrase au contexte pour que le texte soit cohérent, ce lien pouvant être plus ou moins direct.

III.5- Résolution d'un type de métaphore.

Nous avons présenté §I.8, un type de métaphore que nous cherchons à résoudre grâce à notre système. Nous l'avons illustrée par l'exemple suivant: trouver la signification de "soigner" dans "soigner un moteur".

La première tâche consiste à rechercher tous les sens du prédicat et ceux du concept en conflit avec le prédicat. Ce concept est reconnu lors de l'analyse sémantique de la phrase: c'est l'emploi de ce couple qui provoque l'ambiguïté.

Nous allons rechercher les schémas dont le nom contient l'un des deux concepts. Nous sélectionnons aussi les schémas contenant une généralisation du concept complément.

Pour notre exemple, le système connaît deux sens associés à "soigner":

soigner 1: agent sorte-de humain
thème maladie
but animé, partie-de animé
instrument sorte-de instrument

soigner 2: agent sorte-de humain
thème travail

Nous connaissons aussi le sous-graphe suivant pour moteur:

moteur — sorte-de —> objet mécanique — sorte-de —> objet

Nous pouvons donc sélectionner les schémas:

- soigner un travail
- soigner un animé
- ...

Ces schémas sont ceux qui contiennent le prédicat dans leur nom.

- objet mécanique en panne
- réparer un objet mécanique
- vendre un objet
- ...

Ceux-ci contiennent le concept objet, ou un concept plus général.

A partir de ces deux listes, nous constituons des couples contenant un schéma de chaque liste. Un couple est créé lorsque deux schémas appartiennent à une même arborescence dans le graphe modèle. Nous cherchons ainsi les éléments qui dépendent d'une même situation à un niveau plus élevé.

Nous avons déjà donné la signification de ce rapport entre deux schémas. Nous trouvons ainsi deux sens synonymes d'un prédicat dans un certain contexte. Nous pouvons alors trouver le sens du prédicat dans la phrase, et ainsi le nommer.

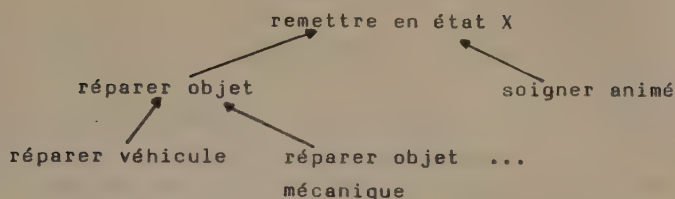
Pour cela, il faut vérifier, ou choisir, le sens ayant un rapport avec le texte.

Lorsque nous constituons les couples, nous les classons selon la généralité du complément. Nous préférons trouver un schéma contenant un complément qui soit le plus proche possible du concept de la phrase.

Pour sélectionner le schéma correct, nous allons chercher un lien entre le schéma pouvant représenter le sens de la phrase, et le contexte. Cette recherche est identique à celle qui est développée lors de la vérification de la cohérence.

Si l'une des hypothèses est confirmée, le schéma sélectionné est celui qui contient le concept objet de la phrase.

Revenons à notre exemple.



Dans notre exemple, nous choisirons "réparer un objet mécanique". Ce schéma constitue un couple avec "soigner un animé"; ils signifient tous deux remettre en état quelque chose ou quelqu'un.

"réparer un objet mécanique" possède aussi un lien avec le contexte, qui est alors constitué des schémas "réparer un véhicule" et "personne être garagiste".

La résolution de ce type de métaphore dépend donc surtout de l'organisation des connaissances. Grâce aux niveaux de précision introduits et à leur structure hiérarchique, nous sommes à même de trouver une synonymie partielle entre concepts.

III.6- Choix du sens d'un mot en cas d'ambiguïté.

L'analyseur de phrases ne peut pas toujours statuer sur le sens des mots. Il est des cas où les seules informations syntaxiques et sémantiques ne suffisent pas à lever les ambiguïtés. Elles ne permettent pas de choisir entre les différents sens d'un terme; plusieurs peuvent convenir.

Cette situation est différente de celle qui est présentée au paragraphe précédent. Le problème ici est de choisir entre les différents sens d'un mot et non de trouver un sens métaphorique du prédicat.

Il faut alors considérer la phrase par rapport à son contexte. Le sens du terme peut être trouvé dans les schémas qui en font partie. Que la phrase contienne un prédicat constituant une entrée vers des schémas ou non, nous cherchons le sens du terme dans les informations contextuelles. En effet le terme est ambigu car nous ne prenons pas suffisamment de connaissances en compte. Ces dernières ne peuvent être trouvées que dans un contexte riche en informations.

Si nous ne trouvons pas de sens au mot traité, cela peut avoir plusieurs significations.

On est en présence d'un changement de thème. Nous pouvons détecter ce changement si le schéma suivant entraîne un changement de thème et permet de trouver le sens du terme.

Il y a toujours la possibilité d'un sens inconnu du système ou d'un emploi métaphorique d'un concept.

Il se peut aussi que les seules informations dont nous disposons ne soient pas suffisantes pour résoudre ce cas.

Mais si nous ne trouvons pas de sens possible au terme de la phrase, nous pouvons déjà éliminer certaines

éventualités.

Lorsque nous cherchons le sens d'un terme, nous partons des différentes interprétations que le système connaît, certaines ayant pu être éliminées par des critères syntaxiques et sémantiques.

La recherche d'un concept dans les schémas se déroule comme lors de la vérification de la cohérence d'une variable. Nous recherchons le concept dans le nom des schémas, dans leurs éléments, qu'ils soient des variables ou des schémas. Dans ce dernier cas, nous prenons le complément en compte. Nous comparons le concept de la phrase avec les concepts décrivant le schéma ou avec les traits de ces concepts.

Si nous trouvons un sens possible, l'analyse de la phrase peut se poursuivre. Si cette dernière fait référence à un schéma, nous pouvons alors chercher le raisonnement permettant de relier la phrase au contexte.

Nous allons donner un exemple de ce traitement. Dans le texte proposé au premier chapitre, se trouve la phrase suivante :

Attendez, je vais prendre un jeu de clefs.

Les mots "jeu" et "clef" possèdent tous deux plusieurs sens; les concepts sont définis de la façon suivante:

prendre : agent animé
thème objet

jeu 1 : action de jouer

2 : ensemble d'objets nécessaires pour jouer

3 : ensemble d'objets

clef 1 : outil

2 : instrument pour ouvrir une porte.

Après l'analyse de la phrase, il reste deux sens possibles pour chacun des mots, le prédicat permettant d'éliminer le premier sens de "jeu".

jeu 2 —————> jouer
 utilité
 |
 —————> objets
 regroupe

clef 1 —————> outil
 sorte de

clef 2 —————> ouvrir
 utilité

jeu 3 —————> objets
 regroupe

Le contexte est composé des schémas:

- objet mécanique en panne
- réparer véhicule

Les concepts ambigus sont reliés par la préposition "de". Le concept introduit par cette préposition permet donc de préciser l'autre concept. Aussi nous commençons par rechercher le sens de "clef".

Dans le schéma "réparer véhicule", il est précisé que l'instrument est un outil, ou plusieurs. Par contre

nous ne trouvons aucun concept compatible avec le second sens de "clef".

Ce concept déterminé, il devient possible de trouver le sens de "jeu" par comparaison des traits sémantiques.

Si nous ne prenions pas en considération la nature de la préposition, la recherche aboutirait aussi, mais après plus de tâtonnements. Rechercher "jeu" dans les schémas ne conduirait pas au choix immédiat de l'une de ses significations. Il faudrait attendre de lever l'ambiguïté sur "clef" pour confirmation.

Les schémas permettent donc de résoudre des cas d'ambiguïtés. Ils apportent plus d'informations sur les concepts, permettent de les relier de nombreuses façons. Mais comme nous n'y avons pas mis toutes les connaissances qu'un humain possède, les schémas peuvent se révéler insuffisants pour résoudre certaines ambiguïtés. Ils permettent alors d'éliminer des possibilités.

Ils peuvent aussi montrer l'ambiguïté de certains concepts. Cela correspond alors à un emploi métaphorique d'un concept qui ne peut être détecté par l'analyste.

Un tel cas est illustré par l'emploi de "outil" dans notre exemple de texte:

on va voir votre outil

"outil" est interprété dans son sens courant, comme un instrument permettant de bricoler, réparer.

L'exploration des schémas du contexte ne permet pas de confirmer ce sens (nous expliquons l'exemple en détail au chapitre suivant). Nous pouvons alors supposer que ce concept est utilisé dans un sens métaphorique, mais nous ne pouvons pas le comprendre uniquement à partir des informations contenues dans les schémas. Dans ce cas la seule connaissance du thème n'est pas suffisante pour trouver le sens d'un concept.

CHAPITRE IV

Explication d'un exemple

L'exemple que nous avons traité dans notre système a été présenté chapitre I (Ex. I.2). Nous le redonnons ici afin de rendre ce chapitre plus facile à suivre.

M. est garagiste.

C. - Monsieur, j'ai un canot automobile dans le port et mon moteur est en panne. Une espèce de pêcheur-matelot-joueur-de-boules sur le quai, m'a dit de m'adresser à vous.

M. - C'est bien gentil de sa part. Mais je ne sais pas si je peux y faire grand chose.
Je m'occupe surtout de voitures, n'est-ce-pas. Je n'ai pas fait d'études spéciales.
Alors si c'est un Diesel ...

C. - Ce n'est pas un Diesel. C'est un Beaudoin.

M. - Ah! les Beaudoin, je les connais un peu. J'en ai soigné un pendant trois ans.
Il était sur le bateau de M. Frère, le Président du Tribunal. Vous le connaissez ?

C. - Non, je ne suis pas d'ici.

M. - Oui, enfin ça ne fait rien. On va voir votre outil.
Ficelle, si on me demande, je suis au port sur le
...
Comment il s'appelle votre bateau ?

C. - Le Pescadou.

M. - Sur le Pescadou. C'est un joli nom.
Attendez, je vais prendre un jeu de clefs.

Nous allons préciser les informations qui nous sont utiles dans ce texte et comment se déroule le processus de compréhension que nous avons développé.

Il a fallu au préalable constituer la base de données. Nous donnons en annexe des exemples de schémas et de la structure qu'ils engendrent.

Les phrases du texte sont transformées en une représentation interne où nous ne précisons que les éléments qui interviennent dans notre analyse.

Nous avons déjà dit que certains prédicats ne sont pas porteurs d'une information déterminante pour notre système. Ces prédicats sont reconnus par le fait qu'ils ne constituent pas des entrées vers des schémas. Ils ne décrivent pas une situation, nous ne les considérons donc pas comme des sujets de discussion possibles. Nous en avons des exemples dans les phrases P1, P2 ...

Ces prédicats permettent d'introduire des personnages ou d'indiquer des déplacements. Ils sont nécessaires

pour vérifier la cohérence des informations entre elles, leur signification. Ils ne donnent pas un point de vue particulier de l'objet.

Nous supposons également que les antécédents des pronoms sont trouvés et que l'on connaît ainsi les concepts génériques qui constituent la phrase, sauf en ce qui concerne les mots ambigus, pour lesquels nous ne connaissons pas le concept associé.

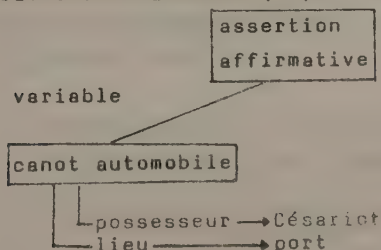
Les différentes questions traitées par cet exemple sont les suivantes:

- recherche du thème principal
- vérification de la cohérence des sujets
- mise à jour du thème actif avec des cas de confirmation, déviation et changement de thème.
- concepts ambigus (P18)
- emploi métaphorique d'un prédicat (P10)
- détection d'un emploi métaphorique d'un concept (P13)

Nous allons maintenant considérer les phrases les unes après les autres et pour chacune, préciser le traitement que nous lui appliquons.

P1: Césariot a un canot automobile dans le port

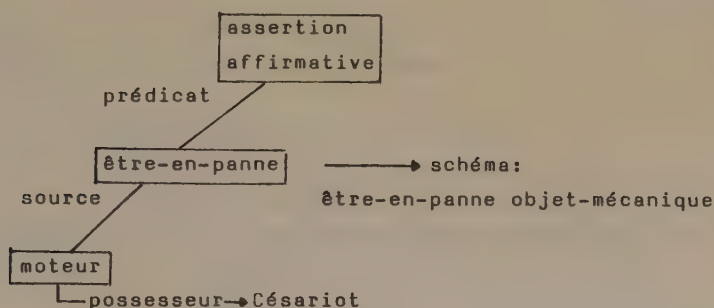
Représentation interne (RI):



Cette phrase étant la première du texte, nous ne pouvons pas trouver un thème principal. Aussi nous conservons la variable. Celle-ci est supposée introduire un élément du thème qui sera, lui, précisé ensuite.

P2: Le moteur de Césariot est en panne

RI:



La phrase permet de sélectionner un schéma. Aussi celui-ci constitue le thème principal (TP).

Nous essayons alors de remplir le schéma avec la variable que nous avons conservée. Nous cherchons un élément du schéma dont la valeur, ou bien coïncide avec le concept "canot-automobile" de la variable, ou bien possède une relation avec ce concept.

Ici nous mettons en évidence le lien "moteur partie de canot automobile". Nous complétons alors l'objet dans le nom du schéma.

P3: Une espèce de pêcheur-matelot-joueur-de-boules sur le quai, m'a dit de m'adresser à vous.

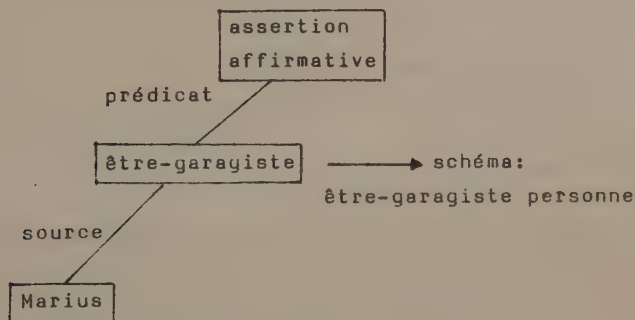
Cette phrase introduit un nouveau personnage, Marius, en expliquant comment Césariot est arrivé jusqu'à lui. Ceci peut être exploité au niveau du traitement des personnages. Ce type de raisonnement n'entre pas dans le processus de compréhension qui nous intéresse ici.

Comme on introduit un nouveau personnage, nous ne retenons de cette phrase que l'information suivante:

Marius est garagiste

C'est l'information qu'un module de traitement du dialogue et des personnages devrait nous communiquer.

RI:



Nous cherchons toujours le thème principal. Le contexte que nous construisons est formé du schéma "être-en-panne objet-mécanique" uniquement. Il n'y a pas de schéma qui le décrit.

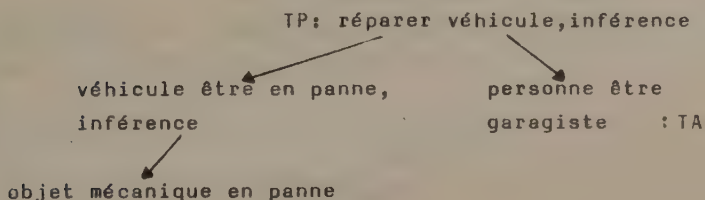
Le chemin trouvé entre le schéma et le contexte est le suivant:

être-garagiste personne → réparer véhicule → ..
 -- → véhicule être-en-panne → objet-mécanique être-en-panne

Nous avons montré au chapitre précédent, §II.1, qu'un tel chemin d'inférences conduit à choisir "réparer véhicule" en tant que thème principal.

Le thème actif (TA) est alors "être garagiste personne".

Nous construisons l'arborescence représentant le déroulement du thème et nous obtenons:

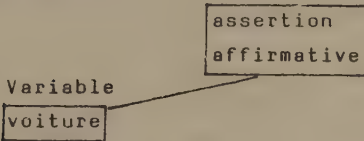


Nous indiquons toujours les noms des schémas modèles car nous travaillons dans le graphe modèle pour rechercher les liens.

Quand nous remplissons un schéma, les référents des différents termes de la phrases ont été trouvés.

P4: Je m'occupe surtout de voitures

RI:



"s'occuper" a un sens trop vague pour constituer une entrée vers un schéma.

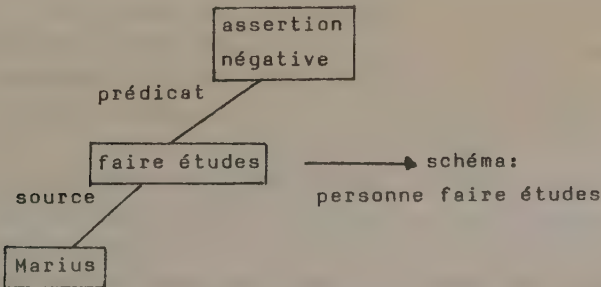
Le contexte est composé de :

- personne être garagiste, TA
- objet-mécanique être-en-panne, ancien TA avant d'aborder le thème actuel
- réparer véhicule, TP

La variable est recherchée dans le contexte. Elle fait partie du schéma "personne être garagiste". C'est donc une confirmation du thème actif.

P5: Je n'ai pas fait d'études spéciales

RI:



Le contexte est le même qu'à la phrase précédente.

Le chemin trouvé est le suivant:

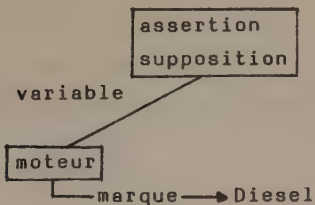
personne faire-études → apprendre matière → connaître
matière → connaître mécanique → personne être-garagiste

Il s'agit donc d'une déviation de thème. On parle d'un point particulier du thème "être garagiste personne". Le thème actif devient "faire des études personne".

L'arborescence représentant le suivi du thème est complétée par ce chemin, à partir du thème développé, en l'occurrence l'ancien thème actif.

P6: Alors si c'est un Diesel ...

RI:



Le contexte est à ce moment:

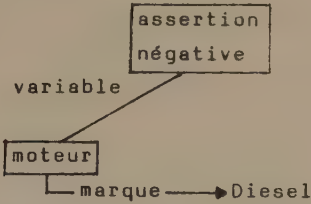
- personne faire-études
- personne être-garagiste
- réparer véhicule

Le traitement relève ici du traitement des référents car nous savons que l'on précise un concept déjà connu. Le suivi du thème permet de préciser le champ de recherche du référent. Seul le thème principal est en liaison avec le concept "moteur", dans "objet mécanique en panne". Ce dernier schéma constitue alors le thème

actif puisqu'il a déjà été instancié. C'est une déviation du thème principal.

P7: Ce n'est pas un Diesel

RI:



Le contexte est :

- objet-mécanique être-en-panne
- réparer véhicule

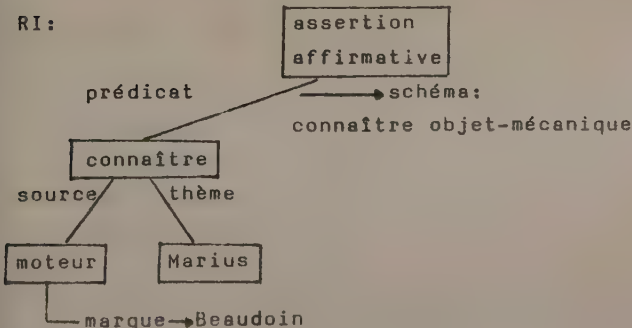
Le traitement est identique à celui de la phrase précédente. Rien n'est modifié au niveau du thème.

- P8: C'est un Beaudoin.

C'est toujours le même processus.

- P9: Ah! les Beaudoin, je les connais un peu.

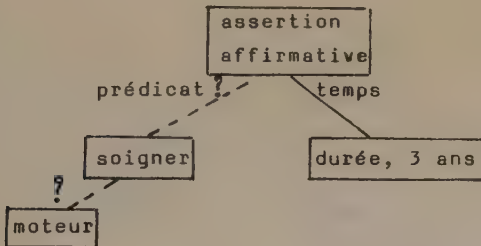
RI:



On relie ce schéma au thème principal par le biais du schéma "être-garagiste personne". Celui-ci ayant déjà été instancié, il devient le thème principal. Le schéma "connaître objet-mécanique" permet alors de le remplir.

- P10 j'en ai soigné un pendant trois ans.

RI



On ne peut déterminer le sens du prédicat car les cas qui le décrivent ne peuvent pas être remplis par "moteur".

Nous avons déjà expliqué le processus mis en place dans un tel cas.

Le schéma équivalent trouvé dans la graphe est donc "réparer objet-mécanique".

Le contexte à ce moment du discours est constitué des schémas:

- personne être garagiste
- objet-mécanique être-en-panne
- réparer véhicule

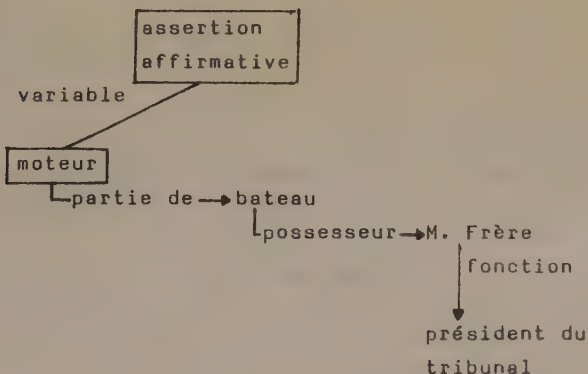
Nous relient alors "réparer objet-mécanique" à

"personne être-garagiste", car il est situé en premier dans le contexte. En effet il y a aussi un lien avec "réparer véhicule".

C'est une confirmation du thème actif.

P11: Il était sur le bateau de M. Frère, le président du tribunal

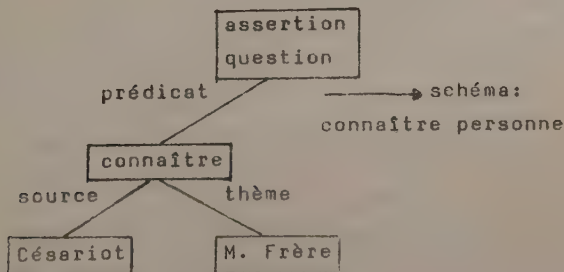
RI:



Cette variable décrit un élément du thème actif (le complément du schéma "réparer objet-mécanique"). C'est une confirmation de ce thème.

P12: Vous le connaissez ?

RI:



Le contexte est:

- personne être garagiste
- objet mécanique en panne
- réparer véhicule

Nous ne trouvons pas de lien entre le schéma appelé par la phrase et le contexte.

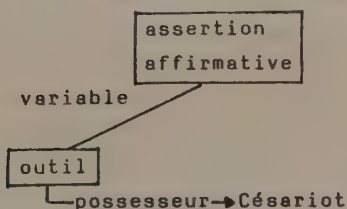
Aussi nous cherchons une relation entre les concepts du schéma et ceux du TA, ainsi que nous l'avons exposé §III.2.

Nous construisons un lien de cohérence grâce au concept "M. Frère", cité dans la phrase précédente et repris dans la dernière phrase par un pronom.

Le thème actif devient "connaître personne".

P13 : On va voir votre outil.

RI:



Nous avons mentionné un problème à ce niveau. En effet le contexte est:

- connaître personne
- personne être garagiste
- réparer véhicule

Nous trouvons la référence à un concept "outil" dans le schéma "réparer véhicule". Nous vérifions si les traits définissant l'emploi de ce concept sont compatibles avec les traits du concept de la phrase. Il se trouve qu'il y a conflit au niveau des possesseurs de cet outil. Nous ne pouvons donc rattacher ce concept au schéma, ni à aucun autre d'ailleurs.

Nous sommes en présence d'un emploi métaphorique d'un concept. Nous ne pouvons pas trouver sa signification exacte. Nous conservons la variable au cas où elle introduit un changement de thème et se rattache au schéma suivant. Dans le cas présent, la variable ne pourra pas lui être reliée.

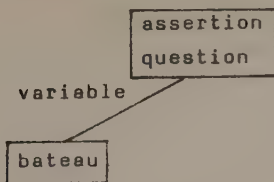
Nous ne modifions pas le thème actif.

-P14: Je suis au port.

Cette phrase précise le personnage de Marius. Cela ne modifie pas le déroulement du thème.

P15: Comment il s'appelle votre bateau?

RI:



Le contexte est:

- connaître personne
- personne être garagiste
- réparer véhicule

La recherche du concept "bateau" dans l'un des schémas permet de trouver "réparer véhicule"; il devient le thème actif.

P16: Le nom est Pescadou

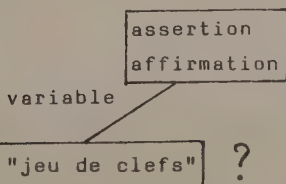
Cela précise le concept de bateau. Rien n'est modifié.

P17: C'est un joli nom

Même chose pour cette phrase.

P18: Attendez, je vais prendre un jeu de clefs

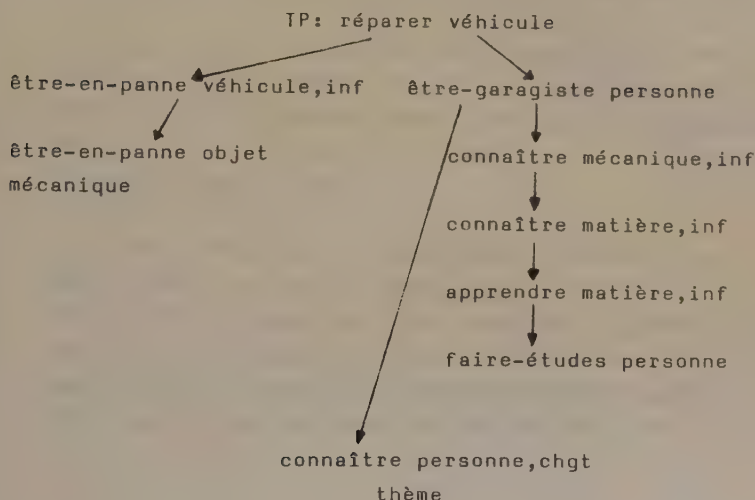
RI:



Nous avons décrit le traitement de cette phrase §III.6. Nous n'y reviendrons pas ici.

Le thème actif reste le même.

REPRESENTATION DU SUIVI DU THEME.



CONCLUSION

Nous avons montré par cet exemple les possibilités et les limites de notre programme. Il serait beaucoup plus efficace s'il disposait d'un traitement des référents, les deux problèmes étant souvent imbriqués. la séparation que nous avons créée est totalement artificielle à certains moments du traitement, notamment lors du traitement de concepts dont on a déjà parlé. A ce niveau, on arrive à la vérification de la cohérence des

informations réelles contenues dans le texte, et ce n'est plus la cohérence des sujets qui est traitée.

Un autre inconvénient peut provenir du nombre de schémas à créer. Mais souvent il s'agit de créer des liens entre situations. Leur description en est faite une seule fois et est accessible à partir de toutes les situations semblables.

Nous avons écrit ce programme en SIMULA. Nous avons déjà signalé que nous n'avons pas approfondi la manipulation des schémas et créé d'autres procédures que celles qui nous sont utiles. Cela aurait pu être fait si nous avions disposé d'un langage adapté à la manipulation des représentations, tel que MERING, Ferber (20).

CONCLUSION

Nous avons présenté dans cette thèse, un système destiné à vérifier la cohérence des textes. La réalisation de cette tâche est indispensable pour résoudre les problèmes suivants :

- suivre le développement du thème
- intégrer chaque phrase dans son contexte afin de pouvoir trouver toute sa signification.

Nous allons situer ce système par rapport aux autres processus intervenant lors de la compréhension d'un texte.

Un tel système n'engendre pas une analyse approfondie du texte. Il en constitue un prélude et se situe après l'analyse syntaxico-sémantique des phrases.

Nous avons vu que ces deux modules ne sont pas indépendants. L'analyseur de phrases a recours à notre système pour résoudre certaines ambiguïtés dans la phrase. Nous avons besoin de la représentation interne des phrases ainsi que de certaines informations syntaxiques dans notre analyse.

Par ailleurs, ce système délimite le champ de recherche des référents et prépare leur résolution. Les

deux modules sont aussi en interaction.

Le système que nous avons réalisé ne prétend pas analyser tous les textes, quelque soit leur structure. La limitation que nous avons imposée réside dans le choix de textes où le thème principal est introduit dès les premières phrases.

Par contre nous n'avons pas limité notre étude à certains sujets uniquement. Ce travail consiste donc à préparer une analyse approfondie des textes étudiés, en délimitant les centres d'intérêts traités. Ainsi, la compréhension du texte pourra s'effectuer à partir d'une représentation globale des thèmes qu'il aborde.

Nous avons insisté sur la mise en oeuvre d'une analyse montante des phrases. Il est très important en effet de limiter les inférences. Nous désirons non seulement limiter ces hypothèses mais aussi limiter la sélection des connaissances à celles qu'il est indispensable d'avoir pour comprendre les phrases et les sujets développés.

Pour réaliser cette analyse, nous avons mis en place une structure de schémas comprenant différents types de liens de descriptions mutuelles et des liens hiérarchiques. Les premiers proviennent de ce qu'un schéma peut être décrit par d'autres schémas. Les seconds correspondent à la constitution de niveaux de précisions pour un sujet donné.

Un problème important réside dans la construction de la base de connaissances. La répartition des connais-

sances dans les différents schémas est tout à fait subjective, ainsi que le choix de ces schémas. Les problèmes qui se posent à nous sont similaires à ceux qui interviennent lors de la constitution d'un réseau sémantique. Nous n'avons pas cherché à résoudre les problèmes spécifiques à un système de représentation: problèmes de cohérence ... Nous proposons ici une structuration et un mode de représentation des connaissances.

La représentation que nous avons mise en place, alliée à notre méthode d'analyse, nous permet de construire un "script" au sens où l'a défini Schank. A partir d'une telle méthode, nous pouvons envisager l'apprentissage de ces "scripts" et ainsi éviter des interprétations quelque peu magiques de phrases telles que :

John alla au restaurant. Il s'assit. Il s'énerva. Il partit.

On imagine aisément plusieurs situations possibles à partir de ces seules informations. Mais les phrases sont comprises et paraphrasées ainsi :

John avait faim. Il décida d'aller au restaurant. Il y alla. Il s'assit sur une chaise. Le serveur ne vint pas à la table. John s'énerva. Il décida qu'il allait partir du restaurant. Il partit.

Les inférences trouvées dépendent essentiellement du déroulement de l'histoire que l'on a entièrement prévue dans le "script" où les cas de déroulements "anormaux" d'un scénario sont aussi répertoriés et

controlés.

Nous prétendons qu'il est inutile de connaître a priori toutes les histoires, mais qu'il est nécessaire de posséder des connaissances fragmentées qui, regroupées par l'analyse du thème, traduisent une histoire dont on vérifie la cohérence.

ANNEXES

- Exemples de schémas . p. 130
- Relations entre schémas p. 133

NOM : réparer véhicule (cas thème)

AGENT : VALDEF: personne (obl)

INSTRUMENT : VALDEF: outil (fac)
 TRAIT(outil,nombre,x)
 TRAIT(outil,possesseur,agent)

CONDITION : N°DIST N°ORDRE
 1 0 * véhicule en panne (obl)
 EGAL(véhicule,compl)
 2 0 * connaître mécanique (obl)
 EGAL(agent,agent)

DESCRIPTION : N°DIST N°ORDRE
 1 1 * trouver panne (obl)
 EGAL(agent,agent)
 2 2 * démonter pièce (fac)
 EGAL(pièce,1.res)
 EGAL(agent,agent)
 EGAL(lieu,lieu)
 3 3 * réparer pièce (obl)
 EGAL(pièce,1.res)
 EGAL(agent,agent)
 4 4 * remonter pièce (fac)
 EGAL(pièce,1.res)
 EGAL(agent,agent)
 SIFAIT(1,obl)

RESULTAT : N°DIST N°ORDRE
 1 0 * véhicule fonctionne (obl)
 EGAL(véhicule,compl)

LIEU : sorte-de local (fac)
 VALDEF: garage

SCHEMA: réparer un véhicule

NOM : être-garagiste personne (cas source)

CONDITION : N°DIST N°ORDRE

1	0	* connaître mécanique (obl)	
		EGAL(agent,compl)	

DESCRIPTION : N°DIST N°ORDRE

1	0	* réparer véhicule (obl)	
		EGAL(agent,compl)	
2	0	* vendre véhicule (fac)	
		EGAL(agent,compl)	
3	0	* connaître objet-mécanique (obl)	
		EGAL(source,compl)	
4	0	* réparer objet-mécanique (obl)	
		EGAL(agent,compl)	

SCHEMA: être-garagiste personne

NOM : être-en-panne objet-mécanique (cas source)

DESCRIPTION : N°DIST N°ORDRE

1	0	pièce	(obl)
		TRAIT(pièce,partie-de,compl)	
		TRAIT(pièce,état,défectueux)	
		TRAIT(pièce,nombre,x)	

SCHEMA: être-en-panne objet-mécanique

NOM : faire-études personne (cas source)

DESCRIPTION : N°DIST N°ORDRE

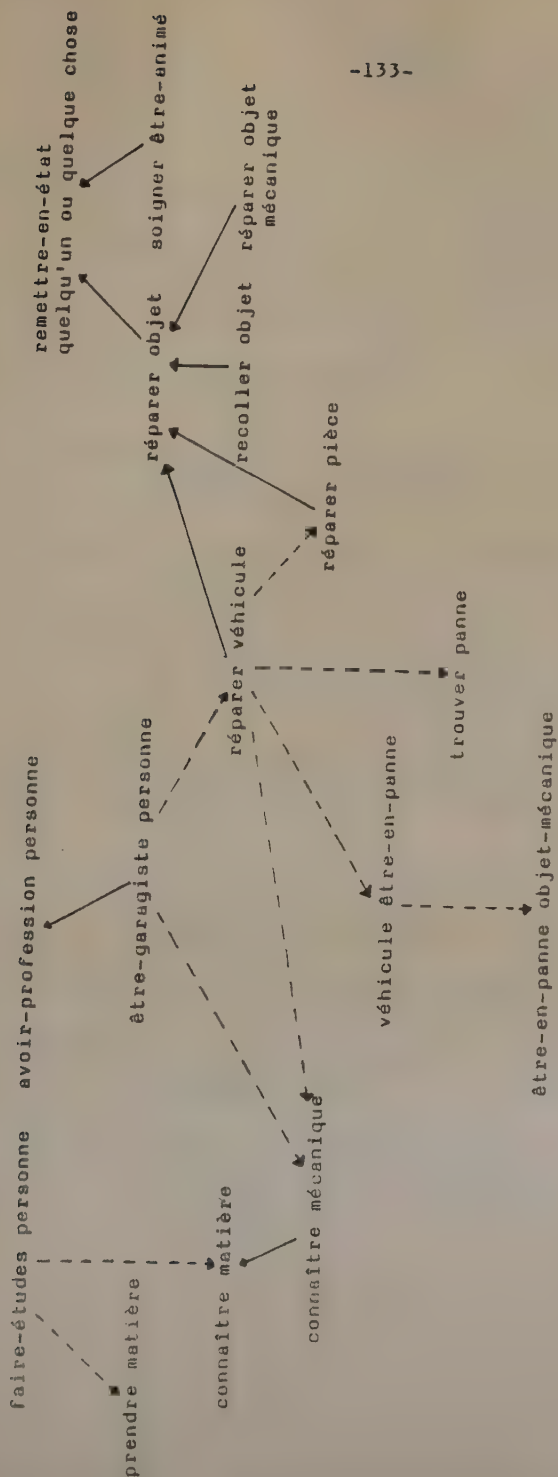
1	1	* s'inscrire dans école	(fac)
		EGAL(agent,compl)	
2	2	* apprendre matière	(obl)
		TRAIT(matière,nombre,x)	
		EGAL(agent,compl)	

LIEU : VALDEF: école

RESULTAT : N°DIST N°ORDRE

1	0	* connaître matière	(fac)
		EGAL(source,compl)	
		EGAL(matière,descr.2.compl)	
2	0	* obtenir diplôme	(fac)
		EGAL(but,compl)	

SCHEMA: faire-études personne



RELATIONS ENTRE SCHEMAS

BIBLIOGRAPHIE

- (1) ANGLIN J.M. Les premiers termes de référence de l'enfant.
La mémoire sémantique. Bulletin de psychologie, spécial annuel 1976.
- (2) ARENS Y. Using language and context in the analysis of text.
7th IJCAI, 1981.
- (3) BERTHELIN J.B., RADY M. Déterminisme et contestation sont dans un bateau
A paraître, ARC, 1984.
- (4) BERTHELIN J.B., SABAH G. Le traitement des personnages.
Congrès AFCET-IRIA, Toulouse, 1979.
- (5) BERWICK R.C. Computational aspect of discourse,
in Computational Models of Discourse ed. by
M. Brady and R.C. Berwick, the MIT Press,
Cambridge, Mass., 1983.
- (6) BOBROW D.G., Winograd T. and the KRL Research Group
Experience with KRL-0. One cycle of a knowledge representation language,
5th IJCAI, 1977.
- (7) BOBROW D.G., Kaplan R.M., Kay M., Norman D.A.,
Thompson H. and Winograd T GUS : a frame
driven dialog system,
Artificial Intelligence, vol 8, n°2, 1977.
- (8) BODEN M. Artificial Intelligence and natural man,
The Harvester Press Ltd, Hassocks, 1977.
- (9) BONNET A. Understanding medical jargon as if it were
a natural language,
6th IJCAI, 1979.
- (10) BULLWINKLE C. (voir aussi Sidner) Levels of complexity in discourse for anaphora disambiguation
and speech act interpretation,
5th IJCAI, 1977.

- (11) BRANSFORD J.D., Mc CARREL N.S., NITSCH K.E.
Contexte, compréhension et flexibilité sémantique : Quelques implications théoriques et méthodologiques,
La mémoire sémantique, Bulletin de psychologie spécial annuel, 1976.
- (12) CHARNIAK E. Towards a model of children's story comprehension
MIT Artificial Intelligence Lab., T.R. 266, 1972.
- (13) CHARNIAK E. Inference and knowledge
in Computational Semantics, North Holland Pub. Co., Amsterdam 1976.
- (14) CHARNIAK E. On the use of framed knowledge in the language comprehension
Artificial Intelligence, vol. 11, n°3, 1978.
- (15) VAN DIJK T.A. The semantics and pragmatics of functional coherence in discourse,
in Pragmatics Ten Years Later, special issue of Versus, ed by J. Boyd & A. Ferrara, 1979.
- (16) DYER M.G., Lehnert W.G. Organization and search processes for narratives
T.R. n°175, Yale, 1980.
- (17) EHRLICH M.F. Apprentissage et mémoire de textes: problèmes théoriques et méthodologiques,
La mémoire sémantique, Bulletin de psychologie spécial annuel, 1976.
- (18) EVENS and al. Lexical semantic relations, a comparative survey,
Linguistic Press, 1980.
- (19) FAHLMAN NEIL, a system for representing and using real-world knowledge,
MIT Press, 1979.
- (20) FERBER J. MERING : un langage d'acteur pour la représentation et la manipulation des connaissances,
Thèse de 3ème cycle, Paris VI, 1983.

- (21) FILLMORE C.J. The case for case
in *Universals in Linguistic Theory*, ed by Bach
& Harms, Holt, Rinehart & Winston Inc., 1968.
- (22) GOLDSTEIN I. and ROBERTS using frames in scheduling
In *Artificial Intelligence, a MIT Perspective*,
vol. 1, ed P.H. Winston and L.H. Brown, the
MIT Press, Cambridge, 1979.
- (23) GRANGER R.H. FOUL-UP: A program that figures out
meanings of words from context,
5th IJCAI, 1977.
- (24) GRAU B. Analyse et représentation d'un texte
d'après le thème du discours
A paraître dans AFCET, 1984.
- (25) GRICE H. Logic and conversation,
in P. Cole and J. Morgan (eds) *Syntax and
Semantics*, vol. 3, 41-58. New-York: Academic
Press, 1975.
- (26) GROSZ B. The representation and use of focus in
dialogue understanding,
Stanford Research Institute, Technical Note
n°151, Menlo Park California, 1977.
- (27) GROSZ B. The representation and use of focus in a
system for understanding dialogs,
5th IJCAI, 1977.
- (28) HALLIDAY System and function in language
Ed Gunther Kress, Oxford Univ. Press, 1976.
- (29) HIRST G. Discourse oriented anaphora resolution,
in *Natural Language Understanding: A review in
Computational Linguistics*, vol. 7, n°2, 1981.
- (30) HOBBS J. Coherence and coreference
Cognitive science, n°3, 1979.
- (31) KINTSCH W., VAN DIJK T.A. Comment on se rappelle et
on résume des histoires,
Langages, n°40, 1975.
- (32) LEHNERT W.G., DYER M.G., JOHNSON P.N., YANG C.J.
and HARLEY S. BORIS : An experiment in in-
depth understanding of narratives, *Artificial
Intelligence*, vol. 20, n°1, 1983.

- (33) LE NY J.F. La sémantique psychologique,
Le psychologue, ed PUF, 1979.
- (34) MAC KEOWN Focus constraints on language generation,
8th IJCAI, 1983.
- (35) MINSKY M. A framework for representing knowledge,
Memo 306, MIT Cambridge, 1974.
- (36) RADY M. L'ambiguïté du langage est-elle la source
du non-déterminisme des procédures de
traitement ?
Thèse de doctorat d'état, Paris VI, 1983.
- (37) REINHART T. The syntactic domain of anaphora,
unpublished Phd Dissertation, Department of
Foreign Literature and Linguistics, MIT,
Cambridge, Massassuchets, 1976.
- (38) REICHMAN R. Conversational coherency,
Cognitive Science, n° 2, 1978.
- (39) SABAH G. Contribution à la compréhension effective
d'un récit,
Thèse de doctorat d'état, Paris VI, 1976.
- (40) SCHANK R.C., ABELSON R.P. Scripts, plans, goals and
understanding,
Lawrence Erlbaum Press, Hillsdale, N.J., 1977.
- (41) SCHANK R.C. Reminding and memory organization: An
introduction to MOPs,
I.R. n°170, Yale, Dpt Computer Science, 1980.
- (42) SIDNER C. Disambiguating references and interpreting
sentence purpose in discourse,
in Artificial Intelligence : A MIT perspective,
vol. 1, ed P.H. Winston and L.H. Brown,
the MIT Press, Cambridge, 1979.
- (43) SIDNER C. Focusing for interpretation of pronouns
AJCL, vol. 7, n°4, 1981.
- (44) SIDNER C. Focusing in the comprehension of definite
anaphora,
in Computational Models of discourse, ed by
M. Brady & R.C. Berwick, the MIT Press,
Cambridge, Massachussets, 1983.

- (45) SIMMONS R.F. Semantic networks: their computation and use for understanding english sentences, in Computer Models of Thought and Language, R. Schank & K.M. Colby (eds), San Francisco: Freeman, 1973.
- (46) SOQUIER S.H. Une interpretation automatique des métaphores. L'exemple des "CHIENG YU", Thèse de 3ème cycle, Paris VI, 1980.
- (47) VILNAT A. Gestion d'un dialogue dans un système de question-réponse, Thèse de 3ème cycle, à paraître.
- (48) WEBBER B.L. So what can we talk about ? in Computational Models of discourse, ed by M. Brady & R.C. Berwick, the MIT Press, Cambridge, Massachussets, 1983.
- (49) WINOGRAD T. Understanding natural language new york, academic press, 1972.
- (50) WOODS W.A., Kaplan R.M. and Nash Webber B., The lunar sciences information system, Final Report, BBN Report 2378, Bolt, Beranek and Newman Inc, Cambridge, Massachussets, 1972.

